PERSONALITY AND RATER LENIENCY: COMPARISON OF BROAD AND NARROW MEASURES OF CONSCIENTIOUSNESS AND AGREEABLENESS

Myranda Grahek, B.S.

Thesis Prepared for the Degree of

MASTER OF SCIENCE

UNIVERSITY OF NORTH TEXAS

May 2007

APPROVED:

Linda Marshall, Major Professor and Chair of the Department of Psychology
Mike Clark, Committee Member
Mark Davis, Committee Member
Sandra L. Terrell, Dean of the Robert B. Toulouse School of Graduate Studies

Grahek, Myranda. *Personality and Rater Leniency: Comparison of Broad and Narrow Measures of Conscientiousness and Agreeableness.* Master of Science (Psychology), May 2007, 81 pp., 11 tables, 9 figures, references, 96 titles.

Performance appraisal ratings provide the basis for numerous employment decisions, including retention, promotion, and salary increases. Thus, understanding the factors affecting the accuracy of these ratings is important to organizations and employees. Leniency, one rater error, is a tendency to assign higher ratings in appraisal than is warranted by actual performance. The proposed study examined how personality factors Agreeableness and Conscientiousness relate to rater leniency. The ability of narrower facets of personality to account for more variance in rater leniency than will the broad factors was also examined. The study used undergraduates' (n = 226) evaluations of instructor performance to test the study's hypotheses. In addition to personality variables, students' social desirability tendency and attitudes toward instructor were predicted to be related to rater leniency. Partial support for the study's hypotheses were found. The Agreeableness factor and three of the corresponding facets (Trust, Altruism and Tender-Mindedness) were positively related to rater leniency as predicted. The hypotheses that the Conscientiousness factor and three of the corresponding facets (Order, Dutifulness, and Deliberation) would be negatively related to rater leniency were not supported. In the current sample the single narrow facet Altruism accounted for more variance in rater leniency than the broad Agreeableness factor. While social desirability did not account for a significant amount of variance in rater leniency, attitude toward instructor was found to have a significant positive relationship accounting for the largest amount of variance in rater leniency.

Copyright 2007

by

Myranda Grahek

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

INTRODUCTION

Performance appraisal ratings provide the basis for numerous employment decisions, including retention, promotion and salary increases. Inaccurate performance appraisal data affect employees' perceptions that ratings are fair and meaningful, and can misrepresent the effectiveness of the organization. Over 75% of the companies sampled by Bretz, Milkovich and Read (1992) reported lenient ratings jeopardized the validity of their performance appraisal systems. Thus, understanding the factors affecting the accuracy of these ratings is important to employees and organizations. Rater leniency is the stable tendency to assign higher ratings in appraisal than is warranted by actual performance. Individual differences likely affect this tendency.

The current study investigated the effects of personality characteristics, attitude toward the instructor and social desirability on rater leniency among students' evaluating instructor performance. It was expected that raters' Agreeableness will be positively related to leniency, whereas Conscientiousness will have a negative association. However, these broad factors are comprised of narrower facets, which are expected to account for more variance in rater leniency. Raters' who hold positive attitude toward their instructor will likely provide more lenient ratings, while individuals high in social desirability are expected to provide less lenient ratings.

The study of job performance is integral to Industrial/Organizational Psychology. The focus of research and practice in industrial (e.g., assessment, selection, training and performance management) and organizational (e.g., leadership, collaborative systems, job satisfaction) psychology is rooted in the ability to predict or improve job performance in some manner. Organizations use job performance information as the basis for employment decisions including salary administration, performance feedback, promotion, terminations, and setting expectations

regarding future performance (Bretz, Milkovich & Read, 1992; Cleveland, Murphy & Williams, 1989). Cleveland et al. (1989) found performance appraisal data were of equal importance for between (e.g., promotion and compensation decisions) and within individual comparisons (e.g., developmental feedback). Thus, the accuracy of performance appraisals is an important issue.

The assessment of job performance occurs through both objective and subjective measures. Objective measures including sales, profits, production, and other quantifiable factors are often unavailable, ambiguous or difficult to assess and compare. An individual's job performance is most frequently measured through subjective measures, such as performance appraisals (Bretz et al., 1992). Subjective evaluations by direct supervisors using a rating scale are the most common (Landy & Farr, 1980; Saal, Downey, & Lahey, 1980), and frequently the only (Bretz et al., 1992) source of performance appraisal information available for an employee. Thus, there are a variety of errors that must be considered.

Subjective measures have several disadvantages, including both random and systematic error, which affect the reliability of the ratings. Random error refers to any nonsystematic deviation of observed scores from true scores and is inherent in all measurement. Rating errors are the undesirable but ubiquitous systematic deviations of observed scores from the true level of performance (Kane, 1994). Of more concern are errors and bias due to the rater. Rater errors are the inaccurate or biased ratings that result from the intentional and inadvertent actions in the recall, processing and evaluation of performance related information (Kane, Bernardin, Villanova & Peryrefitte, 1995).

Common rater errors include halo, central tendency and leniency (Guilford, 1954). The halo effect results from the inability or unwillingness of the rater to discriminate between independent and distinct aspects of an individual's behavior (Cooper, 1981; Saal et al., 1980), the

rater's tendency to attend to a global impression, or a tendency to assign the same rating on different dimensions of performance (Saal et al.). The causes of halo effects can include insufficient motivation and/or knowledge, cognitive distortion, highly correlated and ambiguous performance items (Cooper, 1981). Central tendency errors relate to the differentiation among individuals. A rater's inability to differentiate the performance levels of different ratee's behaviors and shifting scores of all ratees to a common point on the rating scale is referred to as the central tendency error (Kane, 1994; Saal et al., 1980). Additional descriptions include restriction of range and a reluctance to make extreme judgments regarding behaviors (Saal et al.).

Leniency

The current study focused on leniency. Definitions posited by researchers include assigning ratings that are higher than actual performance (Saal et al., 1980; Sharon & Bartlett, 1969), rating level effects (Borman, 1977), and shifting mean ratings above the scale midpoint (Saal et al.). Rater leniency is typically defined as the tendency to consistently shift assigned ratings from the true performance level toward the more favorable end of the scale (Kane, 1994). This definition reflects a general tendency that is stable across time, situations and ratees, and attributable to the individual rater (Hoyt, 2000). Evidence that rater leniency is a stable tendency of the individual (Borman & Hallum, 1991; Kane et al., 1995), supports the examination of individual characteristics including personality and attitudes as sources of the error.

Leniency is especially prevalent in performance appraisal ratings. Typical performance appraisal systems utilize five levels often described on a Likert scale (e.g., 1, below average, to 5, above average). However, the obtained distributions of ratings tend to be skewed, placing 60% to 70% of individuals in the top two levels of performance (Bretz et al., 1992). Thus, organizations typically expect a skewed distribution, generally assumed to be an effect of lenient

ratings (Bretz et al., 1992).

A limitation of past research is inconsistency in the methods used to measure leniency (Saal et al., 1980; Sharon & Bartlett, 1969). Methods have included calculating mean ratings that fall above the scale midpoint or the skewness of the distribution (Sharon & Bartlett, 1969), comparing ratings among raters or across situations (Sharon & Bartlett, 1969), and comparing ratings to mean performance levels (Bartlets & Doverspike, 1997). Midpoint and skewness methods assume that a distribution not normally distributed around the midpoint is inaccurate, which may or may not be the case depending on the characteristic or behavior to be rated. These methods are criticized for not necessarily equating to accuracy or deviation from the true level of performance and can be viewed only as measures of relative leniency among individuals (Murphy & Balzer, 1989).

The most precise or accurate measure compares observed ratings to a known level of true performance or absolute standard of zero leniency (Kane et al., 1995). However, true score estimations are typically available only in the research or laboratory setting. As a result, a compromise between accuracy and practicality must be made. Measures that assess an individual's relative leniency compared to other raters are appropriate when the purpose of the study is to compare individuals or groups of raters. Comparison to midpoint or skewness measures are likely inappropriate unless they are used in combination with other methods.

The true score method has been used by only a few studies (Bartels & Doverspike, 1997; Borman & Hallam, 1991; Jawahar, 2001; Roach & Gupta, 1992). In these studies a true score was estimated based upon an expert's judgment (i.e., researcher or other trained evaluator). Despite the criticism of relative measures of leniency, the majority of studies compared the mean performance ratings between groups or individuals based upon the independent variable of

interest such as research versus administrative purposes (Harris, Smith & Champagne, 1995), type of rating (Thornton, 1980), global versus specific items (Wagner & Griffin, 1997), age, race and gender (Griffeth & Bernadin, 1969; Norton, Gustafson & Foster, 1977; Pulakus, Oppler, White & Borman, 1989), similarity (Turban & Jones, 1988), training (Athey & McIntyre, 1987), beliefs in others (Wexley & Youtz, 1985), lenient ratings from subordinates (Bernardin, 1980), participation in past decisions (Schoorman, 1988), acquaintance (Sundvik & Lindeman, 1998), affect (Antonioni & Park, 2001), self-esteem (Farh & Dobbins, 1989) and rater personality (Yun, Donahue, Dudley & McFarland, 2005). The practicality and usefulness of comparative methods is likely the reason for the prevalence of this technique in research. Additional methods, which were used in a smaller number of studies (Farh & Werbel, 1986; Fox, Caspy & Reisler, 1994; Mount, 1984; Sauser & Pond, 1981; Steele & Ovalle, 1984; Tsui & Berry, 1986; Williams, DeNisi, Meglino & Cafferty, 1986), attempt to identify individual differences that account for a significant amount of the variance in rater leniency (Saal et al., 1980). These methods provide a practical way to study rater leniency that is useful in both applied and research settings.

Sources of Leniency

Landy and Farr (1980) classifies sources of leniency into those attributable to the context, instrument, relationship between the rater and ratee and characteristics of the individual (Landy & Farr, 1980). The context of the rating process refers primarily to the purpose of the ratings, but can also include organizational characteristics (e.g., type, culture, etc.). The type of rating (e.g., self, peer, supervisor, etc.), in addition to the amount of interaction, similarity and affect between the rater and ratee constitute the relationship. The instrument refers to the type of and variations in the rating format used in the performance appraisal (e.g., graphic ratings, forced choice ratings, etc.). Individual characteristics that introduce error include demographic characteristics,

attitudes and personality.

Context Variables

Research examining the relationship between the organizational context and leniency has focused on the setting or purpose of the performance appraisal. Leniency is more prevalent in applied settings where ratings are used to make administrative decisions such as raises and promotions (Bernardin, Orban & Carlyle, 1981; Harris et al., 1995; Jawahar & Williams, 1997; Waldman & Thornton, 1988) or when students' ratings of their peers will affect overall grades (Farh & Werbel, 1986) compared to ratings used for research or to provide developmental feedback to employees. Raters did not appear to provide more lenient ratings when they would be identified (Sharon & Bartlett, 1969), but leniency increased when raters expected their ratings to be validated against other sources (Sharon & Bartlett, 1969; Farh & Werbel, 1986). Leniency appears to be most prevalent when raters are aware their ratings will have real implications.

Research has shown that self-ratings are more prone to leniency than superior or peer ratings (Holzback & Robert, 1980; Mount, 1984; Sundvik & Lindeman, 1998; Thorton, 1980). A review of self-appraisals conducted by Meyer (1980) found that 40% of employees placed themselves in the top 10% of performers. The majority of remaining individuals claimed to be in the top 25% or 50% categories. Only 2% of individuals placed themselves in the below average performance categories. Farh & Dobbins (1989) found that self-ratings were the most lenient when performance dimensions were ambiguous and individuals reported high self-esteem. Research has also compared the degree of leniency in performance ratings conducted by peers, subordinates and supervisors (Antonioni & Park, 2001; Mount, 1984). Research conducted by Mount (1984) found no difference in leniency between supervisory and subordinate ratings. Further research conducted by Antonioni & Park (2001) found employee's ratings of their

supervisors are more lenient than supervisors' ratings of employees and peer ratings.

Instrument Variables

One systematic method of controlling leniency is to select a performance appraisal instrument or variation in how items and scales are formed that decreases leniency. The rating method utilized in most appraisal systems is a graphic rating scale. In an instrument utilizing a typical graphic rating scale, each performance dimension is rated on a 5-point Likert scale. Research has found this method results in more lenient ratings than the forced choice rating scales (Bass, 1956; Sharon and Bartlett, 1969). A forced choice rating scale asks raters to choose between two behaviors, only one of which differentiates high from low performance. The ratee receives one point for the differentiating behavior and no points for the non-differentiating. A mixed standard scale provides three examples of behavior (poor, average, good) for each performance dimension. The rater determines if the individual's performance is better than, equal to or worse than the each behavior. Mixed standard rating methods were also shown to result in less lenient ratings than graphic rating scales (Bass, 1956) and behaviorally anchored rating scales (BARS) (Reardon & Waters, 1979). Rating scales in the BARS method are similar to the graphic rating scale. Performance dimensions are rated on a Likert scale, however a critical incident or behavior is provided for points on the scale. Kingstrom and Bass (1981) conducted a review of 23 studies comparing the BARS method to other rating scales and found no advantages in decreased leniency. However, Tziner (1984) found that when descriptive statements derived from job analysis are used in a BARS format the remaining error is mostly attributed to individual differences.

In addition to assessing the susceptibility of various rating scales to leniency, research has examined the effects of variations to instructions and items. Cautionary instructions noting that

ratings would be compared to other performance measures and altering anchors at the positive end were unsuccessful (Fox et al., 1994). While many efforts to decrease leniency by altering the appraisal format were unsuccessful, items that are clear, specific, relevant to performance and based upon job analysis may decrease leniency in the appraisal process. Wagner and Griffen (1997) found global items were more susceptible to leniency compared to specific items. This finding may be due to the generality of global measures, which is consistent with Farh and Dobbins (1989) findings that ambiguous items increased leniency. Fox, Caspy and Reisler (1994) found leniency was lowest for performance dimensions that were relevant to performance and less lenient for irrelevant dimensions.

Relationship Variables

Numerous variables associated with the relationship between the rater and ratee have been found to affect rater leniency including supervisors' affective response to subordinates, opportunity to observe performance, interpretation of subordinates' self-ratings, demographic similarity and the quality of their working relationship, (Ilgen & Favero, 1985; Judge and Ferris, 1993). Tagger and Brown's (2006) findings supported the positive relationship between leniency and supervisor affect or liking for subordinates. Research has also shown that raters shape their appraisal to fit with their previous decisions regarding the ratee. Schoorman (1988) found individuals who provided input into the hiring decision or agreed with the hiring decision that was made provided lenient ratings, likely to remain consistent with their initial decision. Similar results were found by Williams, DeNisi, Meglino and Cofferty (1986). Research examining leniency in performance appraisal should take into consideration the ongoing relationship between the rater and ratee.

Past interactions including the rater's liking of the individual, may result in more lenient

ratings. Thus, research examining rater leniency should take into account the rater's attitude toward the rater. Consistent with Tagger and Brown (2006) findings that raters' liking of the ratee increases leniency, students who like their class, course material and instructor will be more likely to provide higher ratings to the instructor regardless of actual performance.

Individual Characteristics

Landy and Farr's (1980) review of individual characteristics and performance ratings focused on demographic characteristics including gender, race, age and education. Only one of the eight studies included in this review found gender differences, showing females exhibited greater leniency. A later study found females to be more lenient to males than to females, but no difference in male raters' evaluations (Pulakos et al., 1989). While the studies reviewed by Landy and Farr (1980) had inconsistent findings for age, Griffeth and Bedeian (1989) found a positive correlation. The education of raters had little effect (Landy & Farr, 1980). Race contributed to the rater's perceived similarity to the ratee and resulted in increased leniency (Landy & Farr, 1980), which was supported by Pulakos et al. (1989). Additional studies in perceived similarity supported the positive relationship to more lenient performance ratings (Fox, Ben-Nahum & Yinon, 1989; Turban & Jones, 1988). Overall, differences in rater leniency are unlikely to be related to demographic characteristics, but perceived similarly based upon demographic characteristics appears to increase leniency.

Additional individual characteristics examined in leniency research include self-esteem, self-monitoring, cognitive complexity and leadership style. Individuals low in self-esteem (Farh & Dobbins, 1989; Jawahar & Stone, 1997; Mandell, 1956), self-monitoring (Jawahar, 2001) and cognitive complexity (Bernardin et al., 1981; Schneier, 1977) provided less lenient ratings. Leadership style may also be related to lenient ratings (Landy & Farr, 1980). Production-

orientation, initiating structure (Drory & Ben-Porat, 1980) and consideration (Bernardin, 1980) were positively related to lenient ratings (Landy & Farr, 1980).

A number of individual differences in attitudes and beliefs have been found to affect rater leniency. Bernardin et al. (1981) found individuals who believed others inflated ratings, were more likely to inflate ratings themselves. Raters who believed others could change and were generally altruistic, independent and trustworthy were more lenient in ratings (Wexley & Youtz, 1985). Jawahar (2001) examined the rater's belief that the appraisal system accurately assessed performance and found no difference. Thus, it appears raters' beliefs in others affect leniency, while raters' belief in the appraisal itself does not.

Research examining individual characteristics and leniency has yet to examine variables related to impression management, such as social desirability. Social desirability is a tendency of the individual to present themselves favorably in social interactions (Johnson & Fendrich, 2002). In the organizational setting, performance ratings are often provided in a face-to-face meeting. In these situations raters may provide ratings higher than warranted by actual performance to maintain a favorable opinion of the individual. Individuals high in social desirability may be more likely to provide lenient ratings. When ratings are anonymous raters high in social desirability are likely to provide ratings they believe will be consistent with their peers' ratings. In these circumstances social desirability will be negatively related to leniency.

While all leniency error can be detrimental to the organization resulting in depleted resources for recognizing performance and inadequate documentation for disciplinary action (Bernardin, 1989), leniency that is attributable specifically to the individual is especially important. When significant error is due to individual characteristics, comparisons across raters are difficult and can prevent the organization from comparing individuals' performance across

supervisors or average performance levels across departments. In addition to the characteristics mentioned above, personality variables are likely to affect individuals' tendency to provide lenient ratings in performance appraisals.

Personality

Personality is one way to understand human behavior and experience through individual differences in relatively consistent thoughts, feelings and actions across situations (John & Srivastava, 1999; McCrae & Costa, 1999). Personality traits are believed to describe an individual in terms of general tendencies that are broader than specific behaviors, moods and experiences, but more specific than any universal characteristic. Personality research has provided numerous ways to measure a variety of theoretical conceptualizations at differing levels of specificity. The Big Five model of personality has come to be generally accepted in psychology and used in industrial/organizational studies.

Goldberg (1981) coined the term 'Big Five' to indicate that Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness constituted the core of personality. The model provides a framework for describing personality at its broadest level. The relationship between personality and various workplace behaviors including job performance (Barrick, Mount & Judge, 2001), anti-social behaviors (Lee, Ashton & Shin, 2005; Miller & Lynam, 2001), organizational citizenship behaviors (LePine, Erez & Johnson, 2002), job satisfaction (Judge, Heller & Mount, 2002), motivation (Judge & Ilies, 2002), workplace safety behaviors (Wallace & Vodanovich, 2003), leadership style (Bono & Judge, 2004), entrepreneurial status (Zhao & Seibert, 2006), social influence (Caldwell & Burger, 1997) and rater errors (Bernardin, Cooke, & Villanova, 2000; Yun et al., 2005) have been explained using the Big Five model.

In Costa and McCrae's (1992) model, the factors are typically labeled Extraversion

(talkative, assertive, and energetic), Agreeableness (good-natured, cooperative, and trustful), Conscientiousness (orderly, responsible, dependable), Neuroticism (calm, not neurotic, not easily upset) and Openness (intellectual, imaginative, independent-minded). These five factors are the broadest categorization of personality, which are comprised of the narrower facets (Extraversion – Warmth, Gregariousness, Assertiveness, Activity, Excitement-Seeking, Positive Emotion; Agreeableness -Trust, Straightforwardness, Altruism, Compliance, Modesty, Tender-Mindedness; Conscientiousness – Competence, Order, Dutifulness, Achievement Striving, Self-Discipline, Deliberation; Neuroticism – Anxiety, Angry Hostility, Depression, Self-Consciousness, Impulsiveness, Vulnerability; Openness – Fantasy, Aesthetics, Feelings, Actions, Ideas, Values). Facets are further comprised of covarying traits and specific behavioral tendencies. Costa and McCrae (1985) created and revised the NEO PI™, to measure the Big Three and Big Five, respectively. Neuroticism, Extraversion and Openness to Experience were included in early versions, each comprised of six related facets. Agreeableness and Conscientiousness were added later but did not include the six facet scales until the NEO Personality Inventory, Revised (NEO PI-R) was published (Costa & McCrae, 1999).

Costa & McCrae (1992) described the conceptual basis for the NEO. Openness refers to the degree to which an individual is original, broad-minded, artistic, curious about the world, holds unconventional views and may experience heightened positive and negative emotions. Neuroticism assesses anxiety, anger-hostility, self-consciousness, impulsiveness, and vulnerability to negative affect including depression, fear, embarrassment, anger, guilt and disgust. Extraversion refers to the tendency to be gregarious, warm, assertive and active. Highly extraverted individuals enjoy stimulation, excitement, large groups and gatherings. The interpersonal tendencies of Agreeableness include altruism, courtesy, flexibility, trust,

forgiveness, tolerance, sympathy and helpfulness towards others and cooperativeness. Conscientiousness concerns the control of impulses and the ability to plan, organize and carry out tasks. Individuals high in Conscientiousness are dependable, prudent, scrupulous, punctual, reliable, strong willed and determined.

Only a few studies have examined specific personality variables in relation to rater leniency. Bacon & Navotny (2002) found that Achievement Striving (similar to the NEO's Conscientiousness facet of the same name) was positively related to lenient performance ratings provided by undergraduates to instructors who were lenient graders. Bartels and Doverspike (1997) examined the relationship between leniency in assessment center ratings and the Cattell Sixteen Personality Factors. The factors intelligence, sensitivity and warmth (positively), and second order factor tough poise (negatively) were significantly related to lenient ratings. Cattel's intelligence, sensitivity and warmth factors are similar to the Big Five factors Openness, Extraversion and Agreeableness respectively. The tough poise factor is comparable to low scores on Big Five facets Tender-Mindedness (Agreeableness), warmth (Extraversion), ideas (Openness). These findings support the relationship between various personality characteristics and lenient ratings.

Three studies were found that directly examined the relationship between Big Five factors Agreeableness and Conscientiousness and leniency (Bernardin et al., 2000; Jelley, 2005; Yun et al., 2005). A positive relationship between Agreeableness and lenient ratings was found in all three studies. Two of these supported the hypothesis that Conscientiousness is negatively related to lenient ratings (Bernardin et al., 2000; Yun et al., 2005). Individuals high in Conscientiousness strive for excellence, set difficult goals and maintain high performance standards (Costa & McCrae, 1992). The trustful, sympathetic and cooperative nature of

individuals high in Agreeableness may make them more likely to provide lenient ratings. High Conscientious individuals are less likely to provide lenient ratings as a result of their careful and thorough nature.

In addition to examining the relationship between the broad personality factors measured by the Big Five model (Agreeableness, Conscientiousness, Openness, Neuroticism and Extraversion), Jelley (2005) examined the relationship of leniency to narrow facets of Agreeableness (Altruism, Compliance and Trust), Neuroticism (Angry Hostility) Conscientiousness (Dutifulness) and Extraversion (Gregariousness, Positive Emotions, and Warmth). Only Warmth, a facet of Extraversion, was found to have a significant positive relationship to lenient ratings. In addition to facets of the Big Five factor model, Jelly found narrow facets, Affiliation and Nurturance, measured by Jackson's PRF-E^{TM} were positively related to leniency.

Broad versus Narrow Measures of Personality

The majority of research on the Big Five in the workplace examines the relationship of a group of behaviors (e.g., organizational citizenship behaviors, job performance) to all the Big Five factors and then attempts to explain the significant relationships (Judge, Heller & Mount, 2002; Judge & Ilies, 2002l; Miller & Lynam, 2001; Zhao & Sieber, 2006). Frequently, when no or small relationships are found, researchers suggest the use of narrower measures to explain the criterion of interest (Bono & Judge, 2004; Hogan & Roberts, 1996). Researchers are beginning to propose that rigid adherence to the Big Five is not practical (Borman, 2004; Yun et al., 2005; Paunonen, Rothstien, Jackson, 1999). While the factors are useful for establishing preliminary distinctions and encompassing a wide range of human behavior, they are less useful for predicting specific behaviors (John & Srivastava, 1999). This tradeoff between the breadth of

behaviors that can be explained, and obtaining acceptable validity is often referred to as the bandwidth fidelity problem (Hogan & Roberts, 1996; John & Srivastava, 1999; Paunonen et al., 1999).

While broad personality measures are appropriate predictors of broad criterion such as job performance, narrow measures may be more appropriate predictors of specific behaviors (Ashton, 1998; Ashton, Jackson, Paunonen, Helmes & Rothstein, 1995; Hogan & Roberts, 1996; Jenkins & Griffith, 2004; Murphy & Dzieweczynski, 2005; Paunonen, 1998; Paunonen et al., 1999; Tett, Steele, & Beauregard, 2003). Paunonen (1998) found both types of personality measures contributed additional variance in the behavior criteria, although the additional variance was higher for narrow measures. Ashton et al. (1995) and Jenkins and Griffith (2004) also found lower correlations for broad measures compared to narrow measures. Tett et al. (2003) compared these relationships to job performance in two samples. Low correlations of job performance to broad measures were explained by the differing relationships of narrow measures to the job performance criterion. Conscientiousness, which typically has the highest correlation to job performance, had a low and nonsignificant correlation to productivity (r = .10), a broad measure of job performance. In contrast, narrow measures comprising Conscientiousness, Achievement (r = .75), Cognitive Structure (r = .27), Endurance (r = .26), Play (r = -.74) and Impulsivity (r = -.50) had larger correlations. Accuracy and predictive validity can be maximized by explaining the correlations of narrow personality measures to specific work related behaviors establishing a network of meaningful relationships (Nunnally, 1978; Paunonen et al., 1999) and result in a more thorough understanding.

Current Study

Students' subjective ratings of instructor performance are similar to the subjective ratings

used in performance appraisals in most organizations. The appraisals used in both settings assess various dimensions of performance utilizing a graphic rating scale. Student's ratings of instructor performance provided the dependent variable in the current study.

The current study then compared the relationships of rater leniency to two NEO PI-R factors and the corresponding facet scales, Agreeableness (Trust, Straight-Forwardness, Altruism, Compliance, Modesty, Tender-Mindedness) and Conscientiousness (Competence, Order, Dutifulness, Achievement Striving, Self Discipline, and Deliberation). Table 1 lists the definitions of the factor and facet scales. Consistent with past findings (Bernardin et al., 2000; Jelley, 2005; Yun et al., 2005), it was predicted that personality factors, Agreeableness (positively) and Conscientiousness (negatively) will show a significant relationship to rater leniency.

Hypothesis 1: Rater leniency will be positively related to Agreeableness and negatively related to Conscientiousness.

Three facets of Agreeableness (Trust, Tender-Mindedness and Altruism) were expected to be positively related, one negatively related (Straightforwardness), and two unrelated (Modesty and Compliance) to lenient ratings. The hypothesis for Trust is consistent with Wexley and Youtz's (1985) findings that individuals who believe others are honest, well-intentioned and trustworthy are more likely to provide lenient ratings than individuals low in Trust. Tender-Minded individuals are high in sympathy and concern for others, while Altruism relates to the willingness to act upon that concern. Therefore, Trust, Altruism and Tender-Mindedness are expected to be positively related to rater leniency. It is hypothesized that Straightforwardness will be negatively related to rater leniency. As low scorers are more likely to stretch the truth and use flattery, they may be more likely to provide lenient ratings than individuals high in

Straightforwardness. Compliance may be related to rater leniency when the rater expects to provide feedback, because the individual may attempt to avoid conflict. Individuals in the current study had no expectation of providing feedback to their instructors and the ratings were kept anonymous. Therefore it was predicted that Compliance will be unrelated to rater leniency in this study. Modesty was also hypothesized to have no relationship to rater leniency. The humble and self-effacing nature of modest raters was not expected to affect the leniency of their ratings.

The Conscientiousness facets Order, Dutifulness and Deliberation were expected to be negatively related to rater leniency, while the remaining facets, Competence, Achievement Striving and Self Discipline will be unrelated. Individuals' high in Order and Deliberation tend to be methodical, cautious and deliberate in making their evaluations. Their ratings are more likely to be accurate representations of actual performance and less subject to leniency than individuals low in Order and Deliberation. Dutifulness refers to the tendency to be driven by ethical principals and moral obligations. These individuals are dependable and reliable, and may be more likely to provide accurate or less lenient ratings than individuals low in Dutifulness. While, self-ratings are likely to be affected by an individual's drive and belief in themselves, ratings in others are expected to remain unaffected. There is no reason to believe that individuals' belief in their own capabilities (Competence), ambition and need for success (Achievement Striving) and ability to carry out tasks (Self-Discipline) will be related to lenient ratings.

Hypothesis 2: Individual facets (A1: Trust, A3: Altruism; A6: Tender-Mindedness, positively and A2: Straightforwardness, C2: Order, C3: Dutifulness, C6: Deliberation, negatively) will predict rater leniency.

Hypothesis 3: Individual facets (A4 Compliance, A5: Modesty, C1: Competence, C4:

Achievement Striving, C5: Self-Discipline) will not be significantly related to rater leniency.

Hypothesis 4: Individual facets will provide a significantly stronger overall correlation to rater leniency compared to the broad factors of Agreeableness and Conscientiousness, respectively.

In addition to examination of the personality variables, the current study looked for individual differences in social desirability and attitude toward instructors. The socially desirable response tendency may be apparent in the performance appraisal ratings. It is assumed that the positive relationship between social desirability and rater leniency that would exist in most organizations, where ratings are provided in a face-to-face meeting and affect the ongoing relationship between the rater and ratee, will not occur in this situation. In this study, individuals high in social desirability were expected to provide ratings they believe will be consistent with their peers' ratings. The study also assessed the relationship between rater affect and rater leniency to extend Tagger and Brown's (2006) study showing that liking for the ratee increases rater leniency.

Hypothesis 5a: Social desirability will be negatively related to rater leniency as individuals high in social desirability will provide ratings closer to the mean than those low in social desirability.

Hypothesis 5b: Raters' positive attitude toward the instructor will be positively associated with rater leniency.

METHOD

Participants

Participants (n = 226) were undergraduate students enrolled in psychology courses at the University of North Texas. Students received course credit for volunteering to participate. The sample was 62% female. Participants ranged in age from 18 to 44 (M = 21.174, SD = 3.879). The sample included 23.9% freshmen, 20.6% sophomores, 18.3% juniors, 9.2% seniors, and 4.3% other. Participants had work experience ranging from zero to 29 years (M = 4. 416, SD = 3.760).

Measures

Measures used in the study were an Instructor Evaluation Form and survey consisting of personality scales (NEO Agreeableness and Conscientiousness scales), a social desirability scale, and a scale measuring the student's attitude toward the instructor. In addition, demographic data for gender, age, and years of work experience was collected. Participants were also asked how many times they had evaluated others' performance from 1 (0 to 4 times) to 5 (more than 20 times), how accurate they believed their evaluations are of others from 1 (not at all accurate) to 5 (very accurate) and how comfortable they are in evaluating others performance from 1 (not at all comfortable) to 5 (very comfortable).

Instructor Evaluation Form

The Instructor Evaluation Form utilized for all instructors in the Psychology Department was used in the current study. The form consists of 22 items assessing presentation, class structure and organization, course work and exams, student-instructor relations and course effectiveness; each rated on a graphic rating scale from 1 (greatly below average) to 5 (greatly above average). The form was designed by psychology faculty to assess students' evaluation of

the instructor's quality of teaching and their effective performance in the course.

Rater leniency was calculated as the difference between an individual rater's (student) mean rating to all items on the Instructor Evaluation Form compared to the ratee's (instructor) mean performance rating from all raters. A mean performance rating was calculated for each rater (student) by summing ratings on all items and dividing by the total number of items. A mean rating was calculated for each ratee (instructor) by summing raters' (students') mean ratings and dividing by the total number of ratees. Rater leniency was calculated by subtracting the instructors' mean rating from the student's mean rating. Positive differences indicated lenient ratings. The use of the difference scores allowed for comparison across instructors.

Personality Scales

The Agreeableness and Conscientiousness Factors from the NEO PI-R (Costa & McCrae, 1992) personality inventory were used in this study. Factor and facet scores were calculated for Agreeableness (Trust, Straight-Forwardness, Altruism, Compliance, Modesty, and Tender-Mindedness) and Conscientiousness (Competence, Order, Dutifulness, Achievement Striving, Self Discipline, and Deliberation). Each facet contains eight items, totaling 48 items per factor. Participants respond to a 5-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree) indicating their agreement with each of the items. Factor and facet scores were calculated as the mean of responses to the corresponding items. Coefficient alphas for the facets ranged from .57 to .82 for Agreeableness and .59 to .83 for Conscientiousness in the current sample. Table 2 displays means, standard deviations and Cronbach's alphas for facet and factor scales.

Social Desirability Scale

The Marlowe-Crowne Social Desirability Scale is the primary measure of social desirability used in personality research to assess the tendency to engage in impression

management on self-report scales (Reynolds, 1982). The complete 33-item scale and the short forms contain true false items describing two types of behaviors, those that are socially desirable but unlikely to be demonstrated and common behaviors that are socially undesirable. A high score on the scale is indicative of impression management with a socially desirable response tendency on self-reports. Reynolds short form C will be used in this study. The 13-item form showed the highest reliability ($r_{KR\text{-}20}$ = .76) of three short form versions examined. Reynolds submitted all 33 original items to factor analysis. Items with factor loadings above .40 were included in the initial short form (A), 11 total items. Short forms B and C were developed by including an additional homogeneous item (factor loadings = .39) to increase the internal consistency to an acceptable level. The short form was also shown to have a convergent validity of .96 to the Edwards Social Desirability Scale (Reynolds, 1982). In this study participants responded to each item on a 5-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree) indicating their agreement with each item. An overall score of social desirability was calculated as the mean of responses to the 13 items. The coefficient alpha for the social desirability scale in the current study was .76.

Attitude toward Instructors

Items derived from McCoach and Siegle's (2003) School Attitude Assessment Survey-Revised Attitude toward Teachers subscale was used to assess participants' positive attitude. The Attitude toward Teachers scale is designed to assess students' problems with and hostility towards teachers and significantly differentiate high achievers and underachievers. Scores will be the mean of responses to the 7 items on a 5-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). In the scale development sample, this subscale had a Cronbach's alpha of .89. The Cronbach's alpha in the current study was .90. The original measure referred to

teachers, classes, or school in general. The items (This class is interesting, I relate well to the instructor, I like my instructor, My instructor makes learning interesting, My instructor cares about me, Most of the instructors at this school are good instructors, I like this class) were modified to refer to the specific instructor and course being evaluated.

Procedure

Data were collected from students during the sixth and seventh week of the fall academic semester. Instructors provided the last 30 minutes of class time to collect data from students. Instructors were not present in the classroom while the researcher administered and collected the questionnaires.

Participants were provided an informed consent form briefly summarizing the study, confidentiality of data, and voluntary participation. After signing informed consent forms, participants completed the Instructor Evaluation Form and survey containing the demographic questions, NEO PI-R Agreeableness and Conscientiousness scales, Reynolds short form C Social Desirability scale and Attitude toward Instructor scale. Participants were asked to evaluate the instructor's performance during the course of the semester on the Instructor Evaluation Form. Participants were notified that the evaluations would be used for research purposes only. Completion of the instructor evaluation and survey took participants 15 to 30 minutes.

Pilot Study 1

A pilot study was conducted to assess how altering the social desirability measure from true false response format to a 5-point Likert scale may adversely affect the measure's reliability or distribution. Forty two undergraduate students at the University of North Texas volunteered to participate in the pilot. Two versions of Reynolds Short Form C for the Marlowe-Crowne Social

Desirability Scale were assessed. The first utilized the true false response format and the second a 5-point Likert scale. Participants were provided one of the two versions to complete.

Descriptive statistics (mean, standard deviation, skewness, kurtosis, and Cronbach's alpha) are reported in Table 3. The 5-point Likert scale ($\alpha = .72$) had a coefficient alpha nearly equal to the true false response format ($\alpha = .71$). The histograms displayed in Figure 1 and 2 indicate the 5-point Likert scale was a closer fit to the normal distribution than the true false format. The true false response format also violated the assumption of normality as indicated by a significant *z* score for kurtosis ($z = -1.65$, $p < .05$).

These findings support the use of the 5-point Likert scale response format in the current study. The similar reliability and normal distribution suggest the 5-point Likert scale have equal ability to measure the true variance and individuals' tendency to provide a socially desirable response.

Pilot Study 2

A second pilot study was conducted with undergraduate students enrolled in two summer courses at the University of North Texas ($n = 69$). The sample was 58% female. Participants ranged in age from 20 to 65 ($M = 23.97$, $SD = 5.524$). Work experience for the sample ranged from zero to 40 years ($M = 6.49$, $SD = 5.164$).

Participants in the pilot study completed all measures included in the current study, the Instructor Evaluation Form and survey consisting of personality scales (NEO Agreeableness and Conscientiousness scales), a social desirability scale, and a scale measuring students' attitude toward the instructor. Demographic data for gender, age, and years of work experience were collected. Participants were also asked how many times they had evaluated others' performance and how accurate they believed their evaluations are of others.

Data was collected from students during the final week of the six week summer session. Participants were provided an informed consent form briefly summarizing the study, confidentiality of data, and voluntary participation. The study's survey and instructor evaluation form were administered following the end of the class period after instructors had left the classroom. Participants were informed that their evaluations would be used for research purposes only.

The data analysis described for the current study (see Data Analysis) was also used to examine data collected in the pilot study. The pilot study was administered to 70 participants. Data from one participant was eliminated due to an incomplete survey. The remaining sample (n = 69) does not meet the required sample size determined by the power analysis. The data analysis for the pilot study may lack the power required to identify small relationships as significant, increasing the likelihood of a type II error.

Rater leniency was found to have significant positive relationships to work experience (M = 6.49, SD = 5.16), r = .27, p = .03 and accuracy of evaluations (M = 3.79, SD = .53), r = .27, p = .03. Rater leniency was not significantly correlated with age or number of performance evaluations. Independent samples t-tests found no significant differences between males (M = .05, SD = .54) and females (M = -.02, SD = .62) for rater leniency, $t(66) = .576$, *ns.*

The correlation matrix shown in Table 4 displays the correlations between personality variables (factors and facets), social desirability, attitude toward instructors and rater leniency. These correlations do not support the study's hypotheses that Agreeableness (r = .14, *ns)* and Conscientiousness (r = -.01, *ns)* would be related to rater leniency.

The study hypothesized that three facets (A1: Trust, A3: Altruism and A6: Tender-Mindedness) would be positively related to rater leniency and four facets (A2:

Straightforwardness, C2: Order, C3: Dutifulness and C6: Deliberation) would be negatively related to rater leniency. Results from the pilot sample did not support this hypothesis. A significant positive correlation between Straightforwardness and rater leniency ($r = .26$, $p = .03$) was found. No significant relationships were found among the facets predicted to be related to rater leniency. One facet, Self Discipline ($r = .24$, $p = .04$), hypothesized to have no relationship to rater leniency was found to have a significant positive relationship in the pilot study. Attitude toward instructors ($r = .63$, $p < .001$) was found to have a significant positive relationship with rater leniency in the pilot study. While social desirability ($r = .21$, $p = .09$) was not correlated to rater leniency.

Residual scatter plots using SPSS REGRESSION were examined to assess normality, linearity, and homeoscedasticity in rater leniency as predicted by the Agreeableness and Conscientiousness factors and facets. The plots of standardized residuals and scatter plots of predicted to observed residuals, indicate that these assumptions were met in both instances.

Table 5 displays the results of the hierarchical regression analysis testing the relationship of the Agreeableness and Conscientiousness factors, social desirability and attitude toward instructor to rater leniency. The unstandardized regression coefficients (B), standard error (SE B), standardized regression coefficients (β), and *p*-value for each variable entered are presented. Attitude toward instructor ($\beta = .62$, $p < .001$) was the only variable found to account for a significant amount of the variance in rater leniency.

A second regression analysis was conducted to examine the relationship of the Agreeableness and Conscientiousness facets, attitude toward instructor and social desirability to rater leniency. The results are displayed in Table 6. Again, only the third step in the analysis, attitude toward instructor ($\beta = .57$, $p < .001$) accounted for a significant amount of the variance,

$R^2 = .51$, $F_{inc}(2,54) = 12.78$, $p < .001$.

As data from the pilot study did not provide a sample size large enough to obtain the desired power, the data was examined to determine whether it should be combined with data collected during the current study for final data analysis. Independent sample t-tests were conducted for each of the study's variables, displayed in Table 7. The pilot group ($M = 4.09$, $SD = .71$) provided significantly higher instructor evaluation ratings than the current sample ($M = 3.86$, $SD = .72$), $t(293) = 2.31$, $p = .02$. The two groups significantly differed on 12 of the 16 independent variables.

A number of factors could contribute to differences between the two groups. Students who take courses in the summer session could differ in preferred style and other factors contributing to the evaluations of instructors (e.g., amount of homework assigned, preferred teaching style). One class included in the pilot group was from business department, while all classes in the current study are from the psychology department. Differences between students from the two disciplines could also be a contributing to the disparity between the studies. The differences in class structure (e.g., class size, more time spent with instructor over a shorter period of time during the summer session) could also result in students having different opportunities to observe instructor behavior and form evaluations. Summer session students were also at the end of their term and likely had an accurate idea of the grade they would be receiving, which may impact their instructor evaluations. Students in the current study had only attended 6 to 7 weeks of a 16 week term, having less contact with their instructors and more tentative understanding of their final grade. Due to the large number of factors that could be underlying the differences observed between the two groups, the data from the pilot study will not be included in the final data analysis for the current study.

Data Analysis

Stieger's power analysis procedure was utilized to calculate the required sample size for the proposed study's analysis. The procedure provides an estimate of the required sample size to achieve a desired level of power based upon an expected rho squared in the population, number of variables and alpha. The desired power (1-β) is the probability of finding a relationship that actually exists or probability of rejecting false null hypothesis. The desired level of power in the current study is .80. The second regression analysis in the third step will include 15 variables (rater leniency, social desirability, attitude toward instructors, and the twelve Agreeableness and Conscientiousness facets). Past research examining the relationship between personality factors and rater leniency found significant correlations ranging from .06 to .33 for Agreeableness and .02 to .37 for Conscientiousness (Bernardin et al., 2000; Jelley, 2005; Yun et al., 2005). In the unlikely situation that factors have no shared variance in prediction of rater leniency, an expected multiple r-squared (R^2) would range from .08 to .70. Due to the large variation in values and likelihood of some shared variance an estimate rho squared at the lower end of the range was used to calculate the required sample size ($\rho^2 = .10$). Based upon these values, ($k = 15$), ($1-\beta = .80$), ($\alpha = .05$) and ($\rho^2 = .10$) the required sample size is 180. Approximately 200 participants were recruited for the current study to ensure a sample size large enough to detect the smaller effect size with the desired level of power.

The results of the data analysis will be presented and discussed in further detail in the results section of the current study. First, descriptive statistics were calculated. The range, mean, standard deviation and Cronbach's alpha were calculated for all the scales. In addition, these statistics were reported for the created measure of rater leniency.

Correlations of demographic variables (age, work experience and year in school) and

performance evaluation variables (number, accuracy and comfort) to rater leniency were calculated. A t-test was used to examine differences in rater leniency for male and female raters.

To begin to understand the data, a correlation matrix was constructed consisting of all the study's variables. This included the two broad factors (Agreeableness and Conscientiousness), 12 facet scales (Agreeableness – Trust, Straight-Forwardness, Altruism, Compliance, Modesty, Tender-Mindedness and Conscientiousness – Fantasy, Aesthetics, Feelings, Actions, Ideas, Values), attitude toward instructor, social desirability and rater leniency. Interrelationships among the study's independent variables and relationships to the dependent variable rater leniency were examined.

Prior to the two hierarchical regression analyses testing the hypotheses, examination of data was conducted to test for violations of assumptions relevant to multiple regression analysis. The assumptions that were tested include linearity, normality, and homeoscedasticity, in addition to an identification of possible outliers and collinearity in data. Possible transformation of one or more variables, removal of extreme outliers and/or robust analysis were considered as possible procedures for reducing the effect of violated assumptions on type II error rate, stability of regression coefficients and generalizability of findings. Robust analyses included in S-plus, R or SAS, utilize statistical procedures, such recoding outliers to limit the impact on regression coefficients.

Hierarchical multiple regression procedures were used to test the hypotheses. At each step in regression analyses, significant β and changes in R^2, which indicates that the proportion of variance accounted at each step. Two hierarchical multiple regression analyses were used to examine the additional variance explained by the narrow facet scales over the broad factor scales. Attitude toward instructor and social desirability were entered in the last step of each

analysis to examine for additional variance explained by these variables.

Agreeableness factors were entered into the first step of the first analysis, followed by Conscientiousness factors in the second step. Significant standardized regression coefficients (β), semi partial correlations (*sr*) and multiple r-squared (R^2) support the study's hypothesis that Agreeableness (positively) and Conscientiousness (negatively) relate to rater leniency. A significant change in the total R^2 in the third step would indicate attitude toward instructor and social desirability explain additional variance in rater leniency after accounting for the Agreeableness and Conscientiousness Factors.

The Agreeableness facet scales were entered into the first step of the second regression analysis, followed by Conscientiousness facets in the second step. Standardized regression coefficients (β) were examined for significance to test the hypothesized relationships for the facets. A significant R^2 would indicate that the Agreeableness and Conscientiousness facets predict rater leniency. A significant change in the total R^2 in the third step would indicate attitude toward instructor and social desirability explain additional variance in rater leniency after accounting for the Agreeableness and Conscientiousness Facets.

$$Z^* = (z_{ya} - z_{yb})\sqrt{\frac{N-3}{2-2s_{ya,yb}}}$$

$$z_{ya} = \frac{1}{2}\ln\left(\frac{1+r_{ya}}{1-r_{ya}}\right)$$

$$s_{ya,yb} = \frac{r_{ab}(1-2\overline{r}^2) - \frac{1}{2}(\overline{r}^2)(1-2\overline{r}^2 - r_{ab}^2)}{(1-\overline{r}^2)^2}$$

$$\overline{r} = \frac{r_{ya} + r_{yb}}{2}$$

A larger multiple r-squared than the total R^2 from the first analysis would indicate that a combination of facets may be a better predictor of rater leniency than the broad factors. An additional test was utilized to show that facets are a better predictor of rater leniency than the factors. The equation tests for a significant difference between correlations of adjusted predicted values to observed scores of the two regression analyses (Tabachnick & Fidell, 2001).

RESULTS

The study's survey and instructor evaluation form were administered to 225 undergraduate participants enrolled in one of three classes. Six participants' data were removed due to incomplete surveys or instructor evaluation forms. The remaining sample of 218 participants exceeds the required sample size determined by the power analysis.

Descriptive data and correlations for demographic items (i.e., age, year in school and years of work experience) to rater leniency are displayed in Table 8. Consistent with Landy and Farr's (1980) findings raters' age was not related to leniency in this sample. In addition, no significant relationships were found for year in school or years of work experience. An independent samples t-test was used to test for gender differences in rater leniency. Consistent with Pulakos (1989), there was no significant difference between male (M = .04, SD = .56) and female (M = .02, SD = .54) raters for rater leniency in the current study, $t(214) = .29$, *ns*.

There was no relationship found between rater leniency and the number of times raters reported evaluating other's performance (r = -.01, *ns*) or the rater's comfort with providing evaluations (r = .11, *ns*). The raters' belief in the accuracy of their ratings was found to have a significant positive relationship to rater leniency (r = .22, p = .001). While experience and comfort in providing performance evaluations was not related to leniency, individuals who believed their ratings were accurate provided higher performance ratings.

Hypothesis Testing

Residual scatter plots from the regression analysis assessing the Agreeableness and Conscientiousness factors (see Figure 3 and 4) and the second regression analysis assessing the facets (see Figure 5 and 6) were examined to assess normality, linearity, and homeoscedasticity. The plots of standardized residuals and scatter plots of predicted to observed residuals, indicate that these assumptions were met. However, the scatter plots of standardized to predicted residuals suggested the presence of outliers. Scatter plots of Cook's and Mahalanobis Distance were created to further identify these outliers (see Figure 7 and 8). Two cases (Case 206; Cook's distance = .129 and Case 54; Cook's distance = .104) were found in the facets regression analysis to have Cooke's distances that were twice the distance of all other cases. One of the two cases also had a large Cooke's distance in the factors regression analysis (Case 206; Cook's distance = .052). Examination of univariate histograms found that these two cases had rater leniency values nearly twice that of the nearest case (see Figure 9). Case 206 appeared to be a relative outlier in both regression analyses, while Case 54 appeared to be an outlier in facets regression analyses. Both cases (Case 54 and Case 206) were excluded and regression analyses rerun. Regression equations were also ran in S-Plus using robust analysis to assess the differences in variance accounted for by each model using robust M-estimation method (Factors: R^2 = .26; Facets: R^2 = .30) versus the traditional least square method (Factors: R^2 = .26; Facets: R^2 = .30). The difference in proportion of variance explained by the model was less than .01. Traditional statistical methods using SPSS Regression and excluding the two outliers were used to simplify interpretation.

As the Agreeableness and Conscientiousness factors are linear combinations of the corresponding facets, facets may be highly correlated and multicollinearity could affect the

solution. As shown in the correlation matrix (see Table 10), correlations between facets range from .17 to .56 for Agreeableness facets and .37 to .71 for Conscientiousness facets. Collinearity diagnostics were also examined, although condition indexes ranged from 11.56 to 65.07, no dimension had more than one shared variance proportion above .50. Based upon the recommendations provided by Tabachnick and Fidell (2001) the correlations between facets should have minimal effect on the stability of the results.

The study's first hypothesis stated Agreeableness would be positively related and Conscientiousness negatively related to rater leniency. The correlation matrix shown in Table 9 provides partial support for this hypothesis. Agreeableness was positively related to rater leniency (r = .22, p < .001), while Conscientiousness was found to have a non-significant relationship (r = .13, *ns*).

Hierarchical multiple regression was used to determine the relationship between Agreeableness and Conscientiousness to rater leniency in the current study. The Agreeableness factor was entered into the first step of the equation, followed by Conscientiousness in the second. Results for this regression analysis are displayed in Table 10. The unstandardized regression coefficients (B), standard error (SE B) and standardized regression coefficients (β) for each variable entered, in addition to total R, $R^2$and change in R^2 for each step are presented. Agreeableness accounted for a significant amount of variance in rater leniency in step one, R^2 = .047, $F(1,214) = 10.55$, $p < .001$. However, the addition of Conscientiousness in step two did not result in a significant change in total r-squared, R^2 = .06, $F_{inc}(1,213) = 3.33$, *ns.*

The study's second hypothesis posited three facets (A1: Trust, A3: Altruism and A6: Tender-Mindedness) would be positively related to rater leniency and four facets (A2: Straightforwardness, C2: Order, C3: Dutifulness and C6: Deliberation) would be negatively

related to rater leniency. The three facets predicted to be positively related to rater leniency were found to have significant positive relationships, A1: Trust (r = .18, p < .01), A3: Altruism (r = .32, p < .001) and A6: Tender-Mindedness (r = .17, p = .01). The four facets predicted to be negatively related to rater leniency were found to have no significant relationship, Straightforwardness (r = .09, *ns*), Order (r = .08, *ns*), Dutifulness (r = .13, *ns*), and Deliberation (r = .05, *ns*).

Five facets (A4: Compliance, A5: Modesty, C1: Competence, C4: Achievement Striving and C5: Self Discipline) were expected to have no significant relationship to rater leniency in the third hypothesis. The results of the current study supported this hypothesis. No significant relationships between rater leniency and the five facets were found, A4: Compliance (r = .07, *ns*), A5: Modesty (r = .13, *ns*), C1: Competence (r = .12, *ns*), C4: Achievement Striving (r = .13, *ns*) and C5: Self Discipline (r = .07, *ns*).

A second hierarchical regression was performed to assess the relationships of the facets to rater leniency (see Table 11). The Agreeableness facets were entered into step one, followed by Conscientiousness facets in step two. The Agreeableness facets accounted for a significant amount of variance in rater leniency, R^2 = .12, $F_{inc}(6,209) = 4.80$, p < .001. However, the addition of Conscientiousness facets in step two did not provide a significant addition of explained variance, R^2 = .14, $F_{inc}(6,203) = .84$, *ns*. Examination of individual facets within each step indicates that Altruism (β = .247, p = .001) is the only facet to account for a significant amount of variance in the regression analysis.

The forth hypothesis asserted that the Agreeableness and Conscientiousness facets would provide stronger relationships to rater leniency than the corresponding factors. The total multiple r-squared for Agreeableness factor (R^2 = .05) was lower than then the total multiple r-squared

(R^2. = .12) for the Agreeableness facets, indicating the facets account for more variance in rater leniency than the factor. The same was found for the Conscientiousness factor (R^2 = .06) and facets (R^2 = .14), indicating the Conscientiousness facets account for more variance in rater leniency than the factor also.

In the current study, the individual facet Altruism (β = .316, p = .001) accounted for more variance in the rater leniency than the Agreeableness factor (β = .217, p = .001). In addition, attitude toward instructor and social desirability variables accounted for more additional variance in the analysis examining factors (R^2_{inc} = .20, F_{inc}[2,211] = 27.83, p < .001) than in the facets analysis (R^2_{inc} = .16, F_{inc}[2,201] = 23.46, p < .001).

The equation provided by Tabachnick and Fidell (2001) for comparing sets of predictors was used to determine the ability of facets to better predict rater leniency than the corresponding factors. The difference between predictors was not significant when attitude toward instructors and social desirability were included in the regression analyses (Z* = .032, *ns*). However, when the regressions were recalculated without these variables in the third step, facets were significantly better at predicting rater leniency than factors, Z* = 1.83, p < .05.

The study's final hypothesis stated social desirability would be negatively related and attitude toward instructor positively related to rater leniency. The correlation matrix, displayed in Table 10, indicates social desirability (r = .08, *ns*) had no association. However, attitude toward instructor (r = .46, p < .001) showed a positive correlation to rater leniency in the current study. Social desirability and attitude toward instructor were entered into the third step of both regression analyses to determine the amount of variance in rater leniency accounted for by the two variables. The addition of social desirability and attitude instructor following the Agreeableness and Conscientiousness factors resulted in a significant change in total r-squared,

$\Delta R^2 = .26$, $F_{inc}(2,211) = 27.83$, $p < .001$). However, contribution of social desirability ($\beta = .43$, *ns*) in the prediction of rater leniency was insignificant. The same was found when attitude toward instructor and social desirability were entered into the third step of the regression analysis assessing the Agreeableness and Conscientiousness facets. Addition of the variables resulted in a significant change in total r-squared, $R^2 = .31$, $F_{inc}(2,201) = 23.46$ $p < .001$. However, social desirability ($\beta = .43$, *ns*) did not account for a significant amount of the variance in rater leniency.

DISCUSSION

Subjective ratings are the primary method of conducting performance appraisals in most organizations (Bretz et al., 1992). Research suggests that a proportion of the variance in ratings is due to rater error and not the individual's true performance (Kane et al., 1995). As performance appraisals have important implications for employees and organizations (e.g., salary increases, promotions, retention), it is important to identify possible sources of rater error. One form of rater error, is rater leniency. Rater leniency is the tendency of an individual to provide ratings higher than warranted by actual performance (Saal et al., 1980; Sharon & Bartlett, 1969). Leniency was perceived to jeopardize the validity of performance appraisals in 75% of the organizations surveyed by Bretz et al. (1992). Past research has examined a variety of potential sources of rater leniency. However, a limited number of studies have focused on the relationship of rater personality to leniency (Bacon & Navotny, 2002; Bartels & Doverspike, 1997; Bernardin et al., 2000; Jelley, 2005; Yun et al., 2005).

The purpose of the current study was to examine the relationship of two NEO PI-R Big Five personality factors, Agreeableness and Conscientiousness, to rater leniency. The study expanded upon past research by testing if the corresponding facet scales of Agreeableness (Trust, Straight-Forwardness, Altruism, Compliance, Modesty, Tender-Mindedness) and Conscientiousness (Competence, Order, Dutifulness, Achievement Striving, Self Discipline, and Deliberation) would account for more variance in leniency than the factors. In addition, the relationship of rater leniency to student's attitude toward the instructor and social desirability tendency were examined. Undergraduates' evaluations of instructor performance were used to test the study's hypotheses.

Agreeableness was found to have a significant positive relationship to rater leniency in

the current sample. The study's findings in regards to Agreeableness were consistent with the findings of past research (Bartels & Doverspike, 1997; Bernardin et al., 2000; Yun et al., 2005). Individuals with responses indicative of a trustful, sympathetic and cooperative nature provided instructor evaluation ratings higher than those scoring lower on the Agreeableness factor (Table 10 & 11).

Examination of the relationships of Agreeableness facets provided additional insight into the relationship between individuals' personality and the tendency to provide lenient ratings. Results from this study suggest that individuals who believe others are honest and well-intentioned (Trust), have high sympathy and concern for others (Tender-Mindedness), and are willing to act upon this concern for others (Altruism) are more likely to provide higher evaluations of others than individuals low in these personality characteristics (Table 10). However, results of the hierarchical regression analysis found that only Altruism accounted for a significant amount of the variance in rater leniency (Table 12). The study hypothesized that one facet of Agreeableness, Straightforwardness, would be negatively related to the dependent variable. Results did not support this hypothesis as no significant relationship to rater leniency was found. As hypothesized, the Agreeableness facets, Modesty and Compliance, were not significantly related to rater leniency.

The size of significant correlations found in past research (Bernardin et al., 2000; Jelley, 2005; Yun et al., 2005) for rater leniency Conscientiousness (.02 to .37) Agreeableness (.06 to .33) has spanned a broad range. While the size of correlations in the current research is small according to typical standards, compared to past research the correlation (r = .22) and regression coefficient (β = .23) indicate a moderate to strong relationship in the current sample. The facets Trust (r = .18) and Tender-mindedness (r = .17) had moderate correlations to rater leniency in

the current study. However, the Altruism (r = .32) facet had a strong significant correlation and regression coefficient (β = .25).

The positive correlation between Trust and rater leniency is consistent with Wexley and Youtz's (1985) findings that individuals who believe others are well-intentioned, honest, and trustworthy provide more lenient evaluations. Trust assesses an individual's beliefs in others, Tender-Mindedness taps their feelings toward others, and Altruism gets at the willingness to act upon those beliefs and feelings. As providing performance evaluations ratings requires action, the willingness to act component of Trust may explain why it was the only facet to account for a significant amount of the variance in rater leniency.

Conscientiousness was hypothesized to be negatively related to rater leniency. Examination of correlations and regression coefficients between the Conscientiousness factor indicate that an individual's tendency to be purposeful, strong-willed, determined, scrupulous, punctual, and reliable does not affect their tendency to provide less lenient ratings. The results of the current study also did not support the hypothesized relationships for the corresponding facets of Conscientiousness (Tables 10 – 12). An individual's tendency to be organized and methodical (Order), driven by their conscience (Dutifulness), and think carefully before acting (Deliberation) did not affect rater leniency in student's evaluations of instructor performance in the current sample. As predicted the facets Competence, Achievement Striving, and Self-Discipline were not found to have significantly related to rater leniency. While the study's findings did not support the relationship between Conscientiousness and leniency as found in (Bernardin et al., 2000; Yun et al., 2005), the results are consistent with the non-significant relationship found by Jelley (2005).

The current study posited that narrow facets would be more accurate and meaningful

predictors of rater leniency than the broader factors. Results of the current study support this hypothesis. The single factor Altruism had a higher correlation and accounted for more variance in rater leniency than the broader Agreeableness factor (Tables 10 – 12). In this study, rater leniency was better predicted by examining the individual's willingness to act upon their concern for others (Altruism) than assessing a broader set of interpersonal tendencies (Agreeableness). In addition, attitudes toward instructor contributed less additional variance in rater leniency after accounting of variance explained by Altruism than when entered following the Agreeableness factor. This finding further supports the hypothesis that narrow facets may be better predictors of rater leniency.

In addition to examining the ability to predict rater leniency with individual differences in personality characteristics, the current study examined how students' who like their class, course material, and instructor, as measured by the attitude toward instructor scale, may provide more lenient ratings. The correlation and regression coefficients found in this study indicate that attitude toward instructors was the largest predictor of variance in instructor evaluations. The relationship between favorable attitudes and leniency is consistent with the findings of previous research conducted by Tagger and Brown (2006).

The final hypothesis asserted that an individual's tendency to present themselves favorably in social interactions would be significantly related to less lenient ratings in instructor evaluations. It was posited that individuals higher in social desirability would provide ratings they believed would be consistent with their peers' ratings resulting in a negative relationship to rater leniency. The current study did not support this hypothesis (Tables 10 – 12).

While social desirability did not have a direct relationship to leniency, it is possible that social desirability could have an interaction effect with the personality variables. An individual

high in Conscientiousness and social desirability may provide ratings consistent with their peers' ratings, while individuals high in Agreeableness and social desirability may be more likely to provide more lenient ratings. Post hoc analyses were conducted to examine these relationships. Individuals' social desirability scores were multiplied by the Conscientiousness scores creating new interaction variable. The same procedure was used to create a social desirability by Agreeableness interaction variable. Correlations to rater leniency were then calculated for each of the interaction variables (SD x C and SD x A). Neither interaction resulted in a significant relationship to rater leniency, C x SD (r = .03, *ns*) and A x SD (r = .07, *ns*).

The results of this study have implications for both research and practice. The ability to predict or improve job performance, which requires valid measurement of performance, is at the center of research in Industrial/Organizational Psychology. Organizations depend upon job performance ratings to inform numerous employment decisions including salary increases, promotions, and retention decisions.

In the current study up to 31 percent of the variance in rater leniency was accounted for by differences in individuals' personality and attitudes. An individual's willingness to act upon their concern for others, Altruism (β = .25, $p < .001$) and favorable attitudes toward the instructor (β = .41, $p < .001$) accounted for the majority of this variance (Table 12). These findings identify potential sources of rater leniency in performance evaluations.

Organizations should be cautious when comparing subjective evaluations across raters. The findings of the current study suggest that subjective evaluation ratings are impacted by individual differences including Agreeableness and Altruism. As such, ratings for the same level of performance may differ across raters as a result of these differences. One implication of differences across raters is employees with a lenient supervisor may be more likely to receive the

benefits of high performance evaluations (e.g. salary increases, promotions) than those with a less lenient supervisor. The impact of these differences increase as organizations compare performance of individuals and groups across departments and the organization, resulting in invalid performance data providing the basis for numerous decisions and an inaccurate description of the organizations overall performance.

The study's findings regarding attitude toward instructor also have implications for the validity of performance evaluation ratings. Based upon this study, individuals who have positive attitudes for their instructor are likely to provide higher ratings than individuals with less positive attitudes. In an organizational setting a supervisor may be more likely to provide higher ratings to individuals they favor than individuals with equal performance but are less well liked.

Organizations could potentially use this information and similar findings from other studies to improve the accuracy and validity of performance ratings. Training for raters could be designed to assist individuals in identifying personality tendencies related to lenient ratings. Communicating the types of rater error, developing awareness of personal tendencies, and steps individuals can take to provide more objective ratings could help to curb the effects on performance evaluations.

The study's findings support a need to identify objective measures or alternative criterion for making the administrative decisions, typically made based upon subjective ratings. When applicable, objective measures of performance (e.g. sales, profit, production) should be used in lieu of or in tandem with subjective ratings. Noticeable differences between subjective and objective measures of performance would provide more information and could be used to identify potential reasons for the disparity. This exploration could provide valuable information related to possible rater error, as well as external causes (e.g., market changes, organizational

change) that could be affecting performance. When objective measures are unavailable or too ambiguous to assess or compare across raters, multi-rater tools or 360 evaluations could provide a viable alternative to a single subjective rating of performance. Multi-rater tools which involve collecting performance ratings from multiple sources (i.e., supervisors, peers, direct reports, others) could reduce the impact of a single lenient rater.

The continued examination of alternative rating formats may also reduce the likelihood of rater error typically found with the use of a graphic rating or Likert scale. As suggested by Bass (1956) and Reardon and Waters (1979), the use of a mixed standard rating method or forced choice rating scale (Bass, 1956, Sharon & Bartlett, 1969) may reduce leniency. Alterations to the rating scale such as utilizing clear and specific items (Farh & Dobbins, 1989; Wagner & Griffen, 1997) deriving anchors for behaviorally anchored rating scales from job analysis (Tziner, 1984) or altering anchors at the positive end of the rating scale may also decrease leniency (Fox et al., 1994).

One possible limitation of the current research is the use of an undergraduate sample. While other studies have used undergraduate samples (Bernardin, 2000; Jelley, 2005; Yun et al., 2005) it is unknown how these findings may generalize to performance evaluations conducted in the organizational setting. Students' instructor evaluations typically have less impact on the instructor than supervisors performance ratings. A student's ratings are one of many provided from the class and are typically not identified, protecting the student's anonymity. Past research has suggested that ratings conducted for research purposes are less subject to leniency than ratings used for administrative purposes. Students in the current sample were assured that the ratings provided would be used for research purposes only (Bernardin et al., 1981; Harris et al., 1995; Jawahar & Williams, 1997; Waldman & Thorton, 1988). The research purpose of the

current study could be limiting the study's ability to identify potential sources of the variance in rater leniency due to a restriction of range in the independent variable. The purpose and anonymity of the student's ratings may also result in less intentional and careful ratings. Past research has shown that students reported putting less effort into ratings when they shared responsibility for the evaluations (Petty, White, Oppler, & Boman, 1989). This may affect the reliability and validity of the ratings provided.

The use of an undergraduate sample may have the most significant impact on the attitude toward instructor variable. Student's relationship to their instructors is different from a supervisor to employee relationship in the organization. Students are less likely to have the close personal relationship that may exist between a supervisor and an employee. This may increase the likelihood that the attitude toward instructor variable is more a reflection of the instructors' actual performance and less a measure of liking independent of actual performance. In this case, the relationship of attitude toward instructor to evaluations is likely inflated ($\beta = .25$ to $.27$). The scale used is not clearly differentiated from teacher performance and should be evaluated with caution. Future research could seek to more clearly differentiate the measure used for attitudes or liking from the performance evaluation.

The attitude toward instructor scale used in the current study may also be interpreted as a global measure of overall satisfaction with class, course material and instructor. Wagner and Griffen (1997) found raters were more lenient for global versus specific performance items. If attitude toward instructor is a global measure of performance and the instructor evaluation form a specific measure, it would be expected that students' responses would be strongly correlated with students providing higher ratings to the attitude toward instructor scale than the instructor evaluation form. A large significant correlation was found between these two scales ($r = .77$, $p <$

.001). Future research could use a similar scale as a global measure of performance to test whether raters are more lenient toward global than specific items.

A number of variables that have less applicability in the organizational setting have been found to affect the students' evaluations of instructors, including motivation to learn, ability, amount learned, and expected grade (Wright & Palmer, 2006). Future research could include these types of variables in their studies to better understand the generalizability of an undergraduate sample to the workplace.

Future research could expand or improve upon these findings by including only those facets identified as showing significant relationships. The current study increases its chance of making a Type I error by comparing all 12 facets to the two factors. This approach may be capitalizing on spurious relationships and chance. Future research could focus the examination on a single comparison of Altruism to Agreeableness or developing a composite variable of Trust, Tender-Mindedness and Altruism and than comparing it to Agreeableness.

The current study used a relative measure of leniency, comparing the individual's rating to the mean rating provided by all students. A more accurate or precise measure of leniency would compare the individual ratings to an objective or estimate of true performance. As in most organizational settings this type of comparison measure was unavailable. Relative measures of leniency as used in this study are appropriate and practical when comparing individuals. Future research could establish a better understanding of the relationships among individual differences and leniency by comparing alternative measures of leniency.

The low scale reliabilities measured by Cronbach's alpha for the NEO Agreeableness and Conscientiousness facets may be affecting the ability to identify the true variance accounted for by these variables. The lower scale reliabilities of the narrow facet scales are one disadvantage of

using these measures to predict differences in leniency or other behaviors. Future research could attempt to develop scales with higher internal consistency. The NEO PI-R facet scales were developed using a combination of rational and factor analytic strategies (Costa & McCrae, 1992). However, further attempts to build upon these efforts and possibly lengthen the scales could improve the ability to reliably measure these traits and effectively identify their relationships to behaviors of interest including leniency. Future research could also identify current established measures of current scales that measure the comparable personality traits.

The current research examined the ability to better predict rater leniency with narrow facets of personality measures than the broader Big Five factors, Agreeableness and Conscientiousness. Unlike past research conducted by Jelley (2005), the findings supported the increased amount of variance accounted for by the facets. Narrow facets provided an increased ability to predict rater leniency as well as provided more specific and easier to interpret linkages. The developers of the NEO-PI R have asserted that differences among facets within a single factor are expected "as it reflects real differences in standing on different but related traits." (Costa & McCrae, 1995). While the facets may be related, it does not follow that they will necessarily have the same relationship, in size and direction, to a specific criterion. While, broad factors may be useful in making general descriptions of an individual, examining specific relationships of the facets will provide valuable information in predicting specific behaviors such as rater error as suggested by researchers (Ashton et al., 1995; Jenkins & Griffith, 2004; Paunonen, 1998). The findings of past, current, and future research will provide information to efforts aimed at identifying individuals for training, development of training programs, and design of evaluation systems that minimize leniency.

Table 1

Definitions of Conscientiousness and Agreeableness Facets (Costa & McCrae, 1992)

Factor/Facet	Description
Agreeableness Factor	Primarily a dimension of interpersonal tendencies, agreeableness assesses the degree to which individuals are altruistic, sympathetic of and eager to help others. In turn, high scorers belief that others will be equally helpful.
Agreeableness Facets	
A1: Trust	Refers to the individuals belief and trust in others. High scorers belief that others are honest and well-intentioned; low scorers tend to be cynical or skeptical and belief others are dishonest or dangerous.
A2: Straightforwardness	Assesses the individual's willingness to stretch the truth and be guarded in expressing true feelings. High scorers are frank, sincere and ingenuous; low scorers may use flattery, craftiness or deception to manipulate others.
A3: Altruism	Assesses the individual's willingness to act upon their concern for others. High scorers are generous, considerate of others and willing to assist those in need of help; low scorers are more self-centered and reluctant to get involved in other's problems.

(table continues)

Table 1 *(continued).*

Factor/Facet	Description
A4: Compliance	Assesses the individual's reaction to interpersonal conflict. High scorers defer to others, avoid displaying aggression, and tend to forgive and forget; low scorers are aggressive, compete rather than cooperate and are not reluctant to display anger.
A5: Modesty	Assess the individual's modesty, but does not imply a lack of self-confidence or self-esteem. High scorers are humble and self-effacing; low scorers believe they are superior and may be considered arrogant or conceited.
A6: Tender-Mindedness	Assesses the individual's sympathy and concern for others. High scorers are compelled by other's needs and emphasize the human side; low scorers are stubborn and less moved by appeals to pity.
Conscientiousness Factor	Assess the individual's tendency to be purposeful, strong-willed, determined, scrupulous, punctual and reliable. High scorers display self-control in planning, organizing and carrying out tasks.
Conscientiousness Facets	
C1: Competence	Assesses the individual's sense that they are capable, sensible, prudent and effective. High scorers believe they are well prepared to deal with life; low scorers have a lower opinion of themselves, believe they are inept or unprepared.

(table continues)

Table 1 *(continued).*

Factor/Facet	Description
C2: Order	Assesses the individual's need for organization and order. High scorers are well organized, methodical, and prefer things to be kept neat and tidy; low scorers are unable to get organized and unmethodical.
C3: Dutifulness	Assess the degree to which the individual is driven by their conscience. High scorers adhere to their ethical principles and work to meet their moral obligations; low scorers are more casual regarding their ethics and morals and may be undependable or unreliable.
C4: Achievement Striving	Assess the degree to which the individual is driven by success. High scorers maintain high aspirations, work hard to achieve goals, are diligent and purposeful in the actions; low scorers are lackadaisical, lack ambition and appear aimless.
C5: Self-Discipline	Assesses the individual's ability to begin and carry out tasks despite distractions. High scorers are self-motivated; low scorers procrastinate and are easily discouraged.
C6: Deliberation	Assess the individual's tendency to think carefully before acting. High scorers are cautious and deliberate; low scorers act hastily and speak or act before considering the consequences.

Table 2

Means, Standard Deviations and Cronbach's Alphas for Scales

Variable	M	SD	α
A1: Trust	3.27	.64	.82
A2: Straightforwardness	3.41	.62	.75
A3: Altruism	4.01	.47	.69
A4: Compliance	3.11	.62	.72
A5: Modesty	3.44	.58	.76
A6: Tender-Mindedness	3.51	.47	.57
Agreeableness	3.46	.40	.90
C1: Competence	3.68	.44	.59
C2: Order	3.27	.64	.77
C3: Dutifulness	3.69	.52	.66
C4: Achievement Striving	3.46	.59	.77
C5: Self Discipline	3.29	.67	.83
C6: Deliberation	3.17	.57	.76
Conscientiousness	3.43	.45	.92
Social Desirability	3.01	.52	.76
Attitude toward Instructors	3.56	.88	.90
Teacher Evaluation	3.88	.69	.95

Table 3

Descriptive Statistics for Alternative Versions of the Marlowe-Crowne Social Desirability Scale Short Form C

Response Format	*N*	*M*	*SD*	Skewness	z_s	Kurtosis	z_k	α
True False	21	0.41	.41	-.15	-1.10	-1.60	-1.65*	.71
5-point Likert	21	3.14	.49	-.55	.31	-.08	-.08	.72

Table 4

Intercorrelations between Variables in Pilot Study

Variable	1	2	3	4	5	6	7	8
1. Rater Leniency								
2. Agreeableness	.14							
3. Trust	-.03	.42**						
4. Straightforwardness	.26*	.40**	-.15					
5. Altruism	.12	.15	.01	.01				
6. Compliance	-.14	.60**	.11	.08	-.15			
7. Modesty	-.03	.35**	.09	.05	-.29	.09		
8. Tender mindedness	.22	.34**	-.02	.04	-.11	-.04	-.10	
9. Conscientiousness	-.01	-.14	-.13	-.03	.14	-.31*	-.08	.16
10. Competence	-.15	.16	.15	.15	.05	-.05	.17	-.04
11. Order	.11	.19	.18	-.02	.03	.09	.29*	-.13
12. Dutifulness	.07	.05	.01	-.03	.20	-.08	-.18	.23
13. Achievement Striving	-.18	-.35	-.17	-.01	-.12	-.35**	-.16	.08
14. Self Discipline	.24*	-.12	-.14	.02	-.05	-.15	-.07	.15
15. Deliberation	-.06	-.26*	-.32**	-.16	.07	-.06	-.09	-.09
16. Social Desirability	.21	-.08	.01	.01	.01	-.31**	-.06	.25*
17. Attitude Toward Instructor	.63*	.24*	.05	.25*	.18	-.06	-.14	.34**

(table continues)

Table 4 *(continued).*

Variable	9	10	11	12	13	14	15	16
1. Rater Leniency								
2. Agreeableness								
3. Trust								
4. Straightforwardness								
5. Altruism								
6. Compliance								
7. Modesty								
8. Tender mindedness								
9. Conscientiousness								
10. Competence	.15							
11. Order	.12	-.12						
12. Dutifulness	.53**	-.22	-.24*					
13. Achievement Striving	.39**	.21	-.32**	.03				
14. Self Discipline	.41**	-.16	.06	.02	-.07			
15. Deliberation	.27*	-.38**	.17	-.12	-.17	.22		
16. Social Desirability	.37**	.01	-.21	.46**	.12	.21	-.09	
17. Attitude Toward Instructor	.06	-.16	.05	.17	.01	.08	-.11	.27*

Table 5

Pilot Study: Summary of Hierarchical Regression Analysis for Variables (Conscientiousness and Agreeableness Factors, Social Desirability and Attitude toward Instructor) Predicting Rater Leniency

Variable	*R*	R^2	F_{inc}	*B*	*SE B*	β	*partial*	*Part*
Step 1:	.14	.02	1.34					
Agreeableness				.65	.56	0.14	.14	.14
Step 2:	.14	.02	.02					
Agreeableness				.66	.57	0.14	.14	.14
Conscientiousness				.08	.61	0.02	.02	.02
Step 3:	.63	.40**	20.19**					
Agreeableness				-.05	.47	-0.01	-.01	-.01
Conscientiousness				-.31	.52	-0.06	-.08	-.06
Social Desirability				.07	.13	-0.06	.07	.06
Attitude Toward Instructor				.50	.09	0.62**	.59	.57

*$p < .05$
**$p < .01$

Table 6

Pilot Study: Summary of Hierarchical Regression Analysis for Variables (Conscientiousness and Agreeableness Facets, Social Desirability and Attitude toward Instructor) Predicting Rater Leniency

	Variable	*R*	R^2	F_{inc}	*B*	*SE B*	β	*partial*	*Part*
Step 1		.39	.15	1.80					
	Trust				.06	.24	.03	.03	.03
	Straightforwardness				.56	.25	.27*	.28	.26
	Altruism				.24	.25	.12	.12	.12
	Compliance				-.18	.16	-.13	-.14	-.13
	Modesty				.04	.24	.02	.02	.02
	Tender-Mindedness				.38	.21	.22	.22	.21
Step 2		.53	.28	1.72					
	Trust				-.04	.27	-.02	-.02	-.02
	Straightforwardness				.55	.25	.26	.28	.25
	Altruism				.23	.26	.12*	.12	.10
	Compliance				-.28	.17	-.21	-.21	-.18
	Modesty				-.02	.25	-.10	-.01	-.01
	Tender-Mindedness				.34	.21	.19	.21	.18
	Competence				-.32	.23	-.20	-.19	-.16
	Order				.16	.26	.08	.08	.07
	Dutifulness				-.05	.14	-.05	-.05	-.04
	Achievement Striving				-.35	.23	-.21	-.20	-.17
	Self Discipline				.39	.27	.18	.19	.16
	Deliberation				.34	.24	-.20	-.19	-.16

(table continues)

Table 6 *(continued).*

	Variable	*R*	R^2	F_{inc}	*B*	*SE B*	β	*partial*	*Part*
Step 3		.72	.51**	12.78**					
	Trust				-.16	.22	-.08	-.10	-.07
	Straightforwardness				.22	.22	.11	.13	.09
	Altruism				-.10	.23	-.01	-.01	-.01
	Compliance				-.25	.15	-.18	-.22	-.15
	Modesty				.04	.21	.02	.02	.02
	Tender-Mindedness				.01	.19	.01	.01	.01
	Competence				-.09	.20	-.05	-.06	-.04
	Order				.05	.22	.03	.03	.02
	Dutifulness				-.06	.13	-.05	-.06	-.04
	Achievement Striving				-.41	.19	-.25*	-.28	-.21
	Self Discipline				.34	.23	.15	.20	.14
	Deliberation				-.21	.21	-.12	-.14	.10
	Social Desirability				.01	.14	.01	.01	.01
	Attitude Toward Instructor				.46	.10	.57**	.55	.46

*$p < .05$
**$p < .01$

Table 7

Mean, Standard Deviation, and Independent Sample T-test for Pilot and Current Study

	Pilot Study		Current Study		
	M	*SD*	*M*	*SD*	*t*
Teacher Evaluation	4.09	.71	3.86	.72	2.31*
Agreeableness	3.10	.13	3.03	.14	3.64**
Trust	3.05	.29	2.97	.34	1.76
Straightforwardness	3.09	.28	2.82	.32	6.26**
Altruism	3.10	.30	3.20	.26	-2.55*
Compliance	3.06	.44	2.78	.37	5.31**
Modesty	3.09	.31	3.3	.34	-4.56**
Tender-Mindedness	3.18	.34	3.08	.32	2.23*
Conscientiousness	3.07	.12	2.93	.13	8.04**
Competence	3.15	.36	2.70	.33	9.70**
Order	2.75	.30	2.82	.30	-1.69
Dutifulness	3.32	.56	2.80	.31	9.88**
Achievement Striving	3.20	.36	3.03	.31	3.97**
Self Discipline	3.03	.27	3.02	.34	.23
Deliberation	2.96	.34	3.21	.38	-5.01**
Social Desirability	2.94	.51	3.01	.52	-1.07
Attitude Toward Instructor	3.77	.72	3.55	.89	2.07*

Table 8

Means, Standard Deviations and Correlations to Rater Leniency

Variable	M	SD	*r*
Age	21.18	3.89	-.05
Year in School	2.31	1.62	-.04
Years of Work Experience	4.43	3.77	-.06

Note. No significant relationships.

Table 9

Intercorrelations between Variables

Variable	1	2	3	4	5	6	7	8
1. Rater Leniency								
2. Agreeableness	.22*							
3. Trust	.18**	.75**						
4. Straightforwardness	.09	.74*	.42**					
5. Altruism	.32**	.69*	.50**	.40**				
6. Compliance	.07	.75**	.56**	.43**	.40**			
7. Modesty	.13	.58*	.16*	.43**	.25**	.27**		
8. Tender mindedness	.17*	.69**	.48**	.33**	.46**	.46**	.30**	
9. Conscientiousness	.13	.03	.07	.02	.23**	.01	-.14*	-.05
10. Competence	.12	-.07	.07	-.09	.15*	-.08	-.25**	-.03
11. Order	.08	-.16	-.09	-.10	.04	-.17*	-.15*	-.16*
12. Dutifulness	.13	.15*	.10	.13	.33**	.06	-.02	.08
13. Achievement Striving	.13	-.03	.08	-.05	.20**	-.08	-.20**	-.04
14. Self Discipline	.07	.08	.15*	.03	.22**	.05	-.07	-.02
15. Deliberation	.05	.15*	.03	.17*	.16*	.26**	.00	-.03
16. Social Desirability	.08	.59**	.54**	.48**	.43**	.50**	.21**	.30**
17. Attitude Toward Instructor	.46*	.11	.11	-.00	.14*	.06	.07	.12

(table continues)

Table 9 *(continued).*

Variable	9	10	11	12	13	14	15	16
1. Rater Leniency								
2. Agreeableness								
3. Trust								
4. Straightforwardness								
5. Altruism								
6. Compliance								
7. Modesty								
8. Tender mindedness								
9. Conscientiousness								
10. Competence	.76**							
11. Order	.77**	.47**						
12. Dutifulness	.78**	.56**	.46**					
13. Achievement Striving	.82**	.60**	.53**	.57**				
14. Self Discipline	.88**	.62**	.62**	.63**	.71**			
15. Deliberation	.66**	.38**	.41**	.48**	.37**	.45**		
16. Social Desirability	.31**	.20**	.05	.35**	.16*	.36**	.34**	
17. Attitude Toward Instructor	.02	.06	-.02	.06	.03	.01	-.01	.033

Table 10

Summary of Hierarchical Regression Analysis for Variables (Conscientiousness and Agreeableness Factors, Social Desirability and Attitude toward Instructor) Predicting Rater Leniency

Variable	*R*	R^2	F_{inc}	*B*	*SE B*	β	*partial*	*Part*
Step 1	.22	.05**	10.55**					
Agreeableness				0.30	0.09	0.22**	.22	.22
Step 2	.25	.06**	3.33					
Agreeableness				0.30	0.09	0.21**	.22	.22
Conscientiousness				0.15	0.08	0.12	.12	.12
Step 3	.51	.26**	27.83**					
Agreeableness				0.32	0.11	0.23**	.21	.18
Conscientiousness				0.18	0.08	0.15*	.16	.14
Social Desirability				-0.12	0.08	-0.11	-.10	-.09
Attitude Toward Instructor				0.27	0.04	0.43**	.45	.43

*$p < .05$
**$p < .01$

Table 11

Summary of Hierarchical Regression Analysis for Variables (Conscientiousness and Agreeableness Facets, Social Desirability and Attitude toward Instructor) Predicting Rater Leniency

	Variable	*R*	R^2	F_{inc}	*B*	*SE B*	β	*partial*	*Part*
Step 1		.35	.12**	4.80**					
	Trust				0.08	0.08	0.09	.07	.07
	Straightforwardness				-0.07	0.07	-0.08	-.07	-.06
	Altruism				0.38	0.10	0.32**	.26	.26
	Compliance				-0.10	0.07	-0.12	-.10	-.09
	Modesty				0.08	0.07	0.09	.08	.08
	Tender-Mindedness				0.04	0.10	0.03	.03	.03
Step 2		.38	.12**	.84					
	Trust				0.09	0.08	0.10	.08	.07
	Straightforwardness				-0.07	0.07	-0.08	-.06	-.06
	Altruism				0.34	0.10	0.28**	.23	.22
	Compliance				-0.07	0.08	-0.08	-.06	-.06
	Modesty				0.12	0.07	0.13	.18	.11
	Tender-Mindedness				0.04	0.10	0.03	.03	.02
	Competence				0.12	0.11	0.10	.07	.07
	Order				0.08	0.08	0.10	.07	.07
	Dutifulness				0.05	0.10	0.04	.03	.03
	Achievement Striving				0.08	0.09	0.08	.06	.05
	Self Discipline				-0.16	0.09	-0.19	-.12	-.11
	Deliberation				-0.01	0.08	-0.01	-.01	-.01

(table continues)

Table 11 *(continued).*

	Variable	*R*	R^2	F_{inc}	*B*	*SE B*	β	*partial*	*Part*
Step 3		.55	.31**	23.46**					
	Trust				0.08	0.07	0.09	.07	.06
	Straightforwardness				-0.01	0.07	-0.01	-.01	-.01
	Altruism				0.29	0.09	0.25**	.22	.19
	Compliance				-0.07	0.08	-0.07	-.06	-.05
	Modesty				0.09	0.07	0.09	.09	.08
	Tender-Mindedness				0.01	0.09	0.01	.01	.01
	Competence				0.08	0.10	0.06	.05	.04
	Order				0.08	0.07	0.10	.08	.07
	Dutifulness				0.03	0.09	0.03	.02	.02
	Achievement Striving				0.06	0.09	0.06	.05	.04
	Self Discipline				-0.11	0.09	-0.13	-.09	-.07
	Deliberation				0.01	0.08	0.01	.01	.01
	Social Desirability				-0.06	0.09	-0.06	-.05	-.04
	Attitude Toward Instructor				0.25	0.04	0.41**	.43	.40

*$p < .05$
**$p < .01$

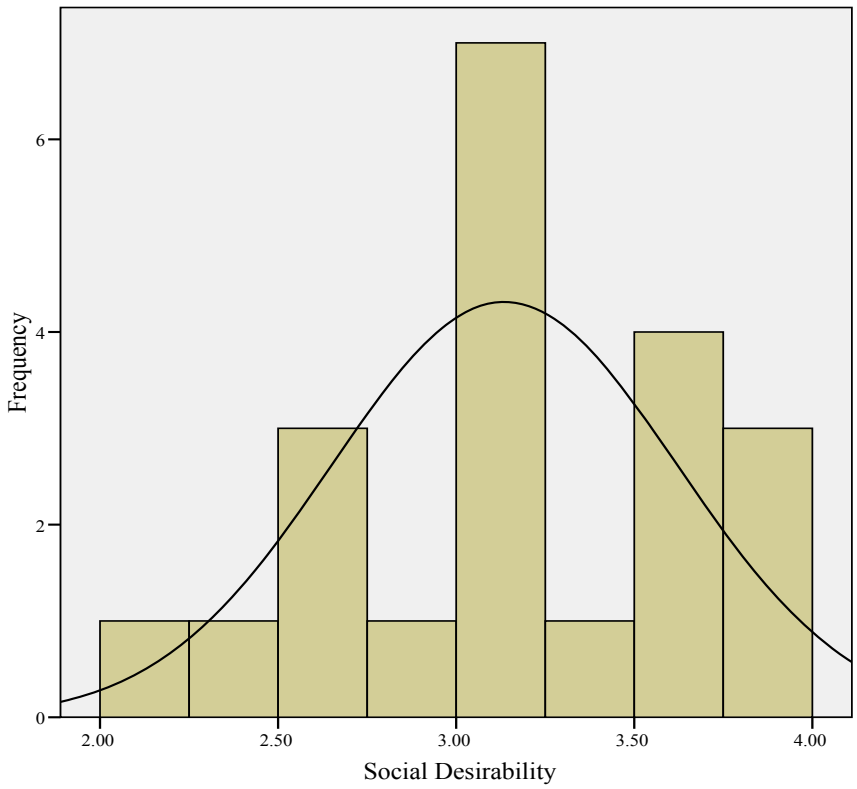

Figure 1. Histogram for social desirability (5-pt Likert scale).

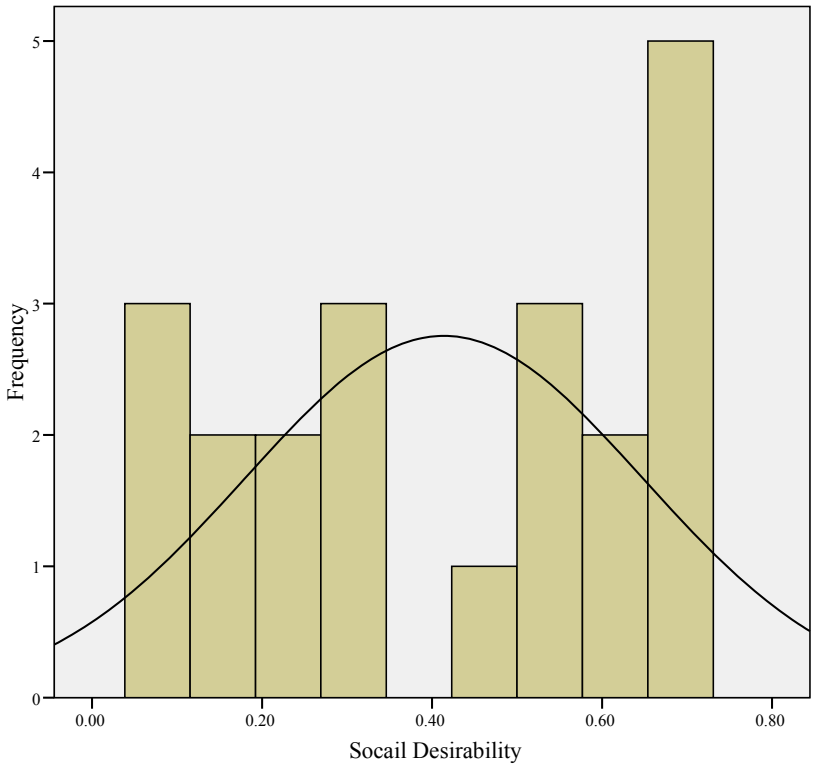

Figure 2. Histogram for social desirability (true/false response format).

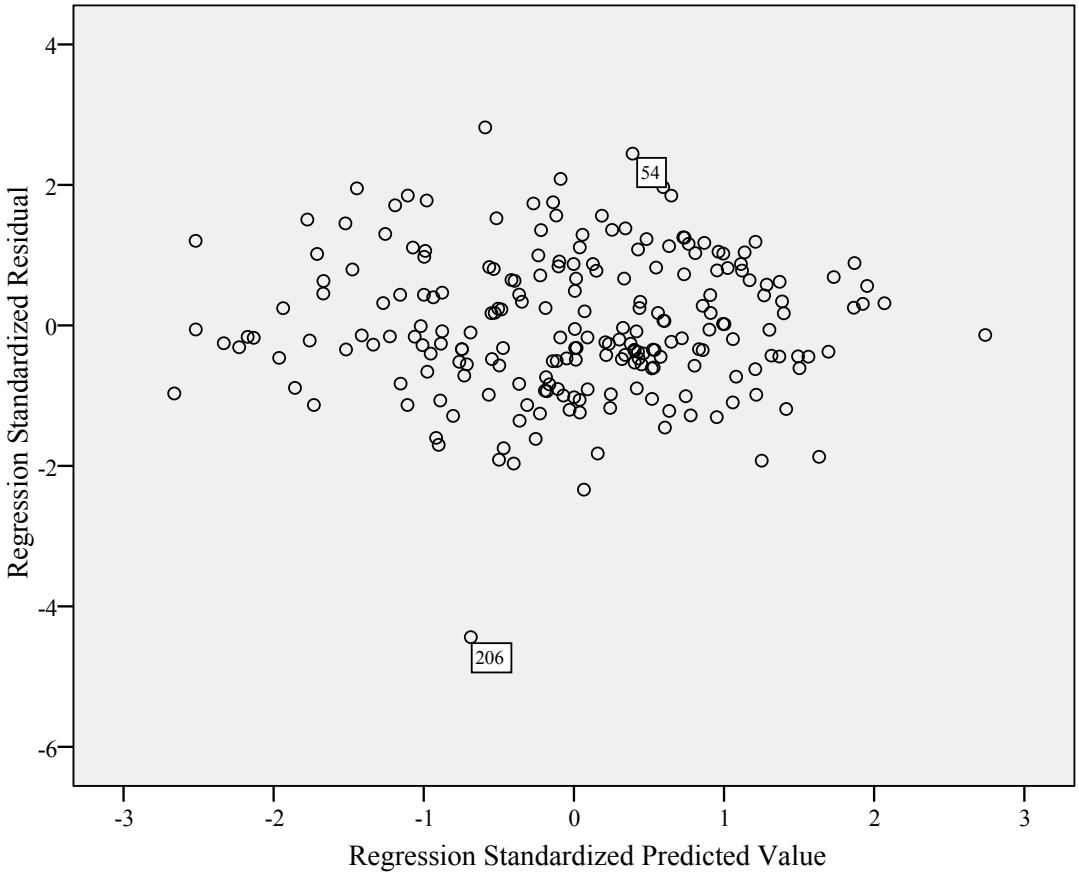

Figure 3. Scatter plot of residuals to predicted values for factors analysis.

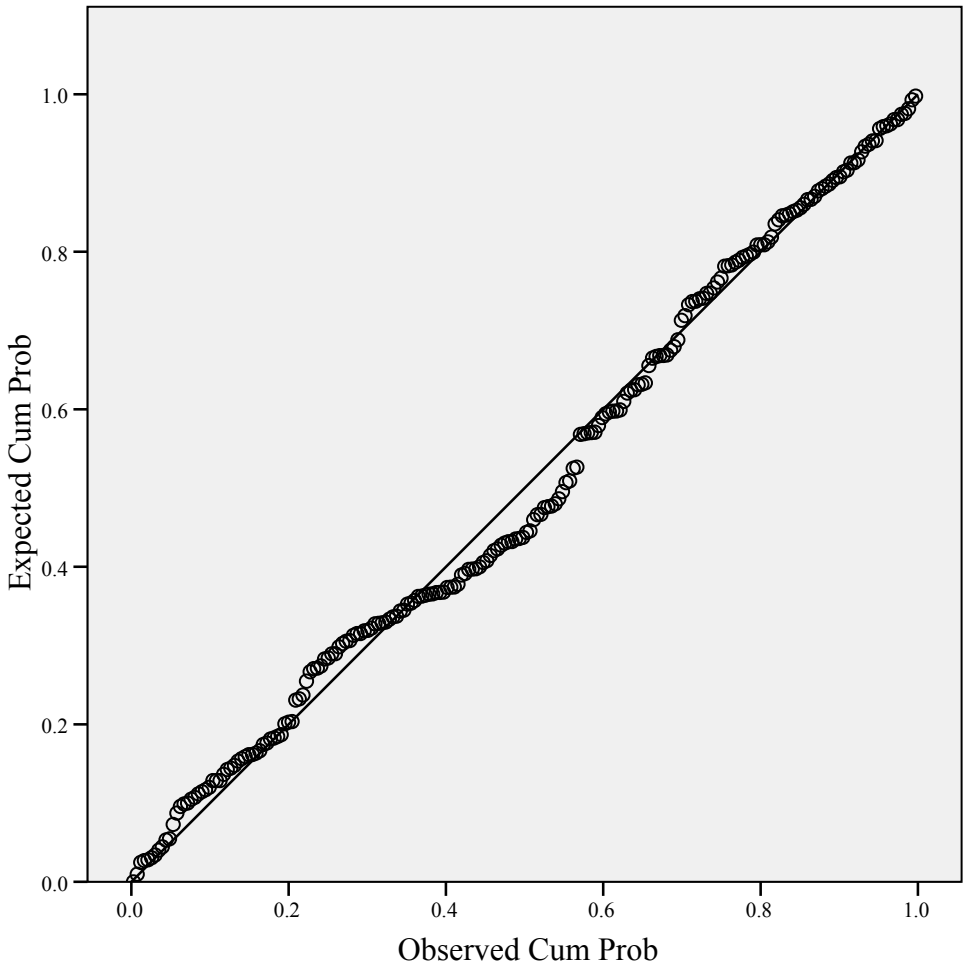

Figure 4. Normal *p-p* plot of regression standardized residuals for facets analysis.

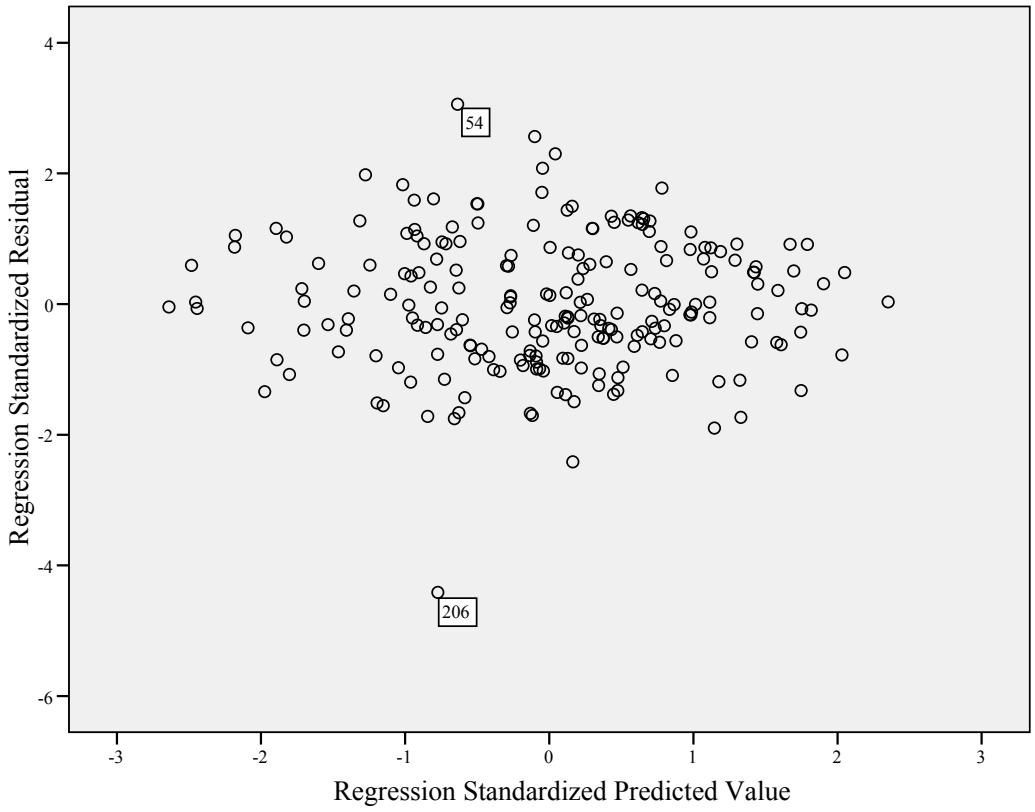

Figure 5. Scatter plot of residuals to predicted values for factors analysis.

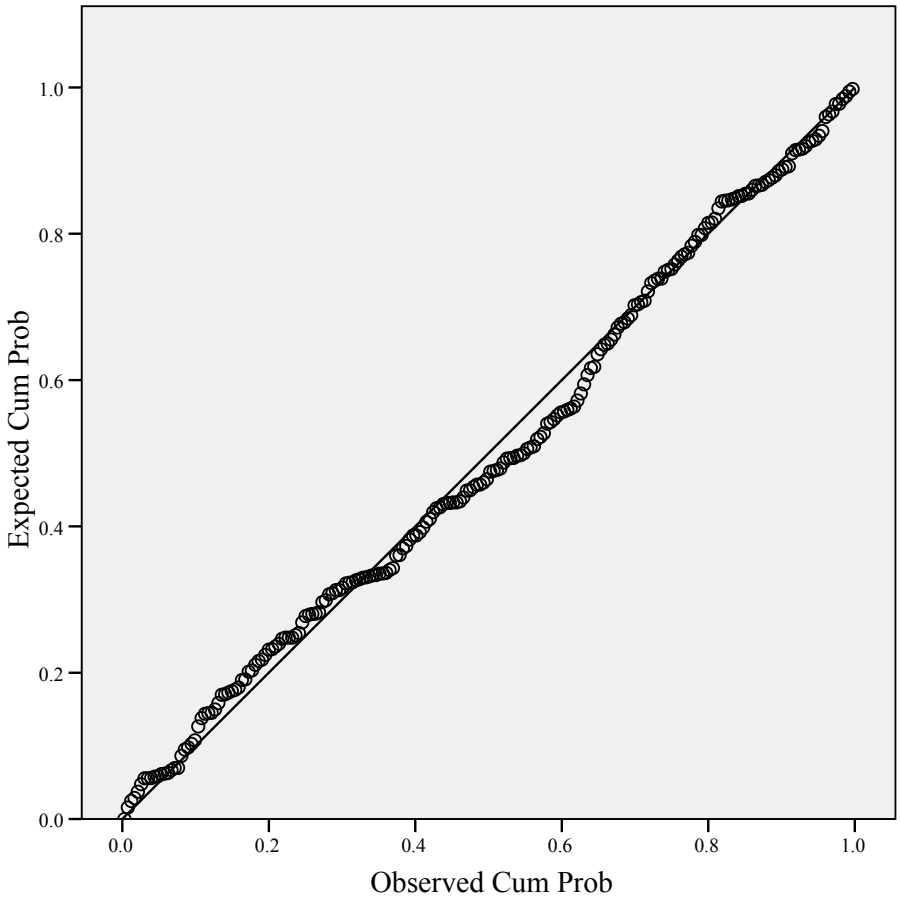

Figure 6. Normal *p-p* plot of regression standardized residuals for facets analysis.

.

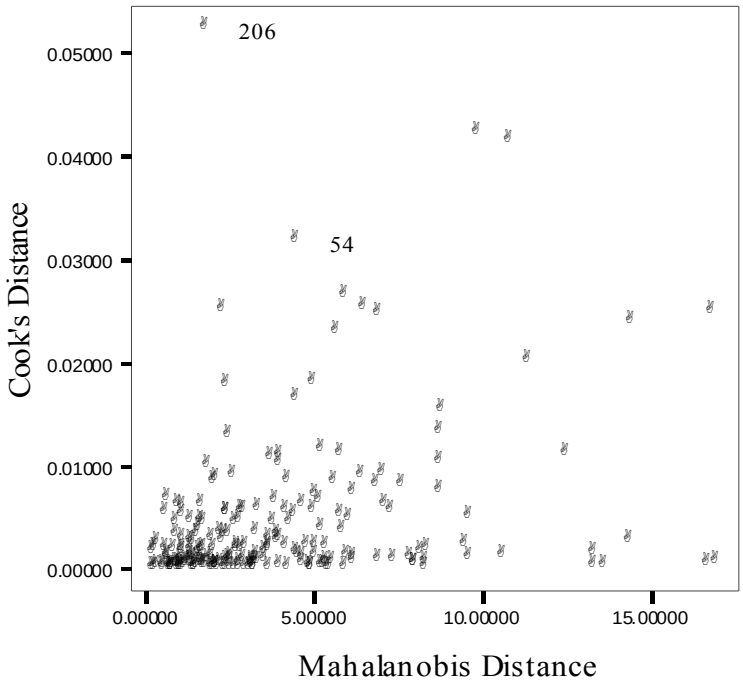

Figure 7. Scatter plot of Cook's to Mahalanobis distance for factors analysis.

.

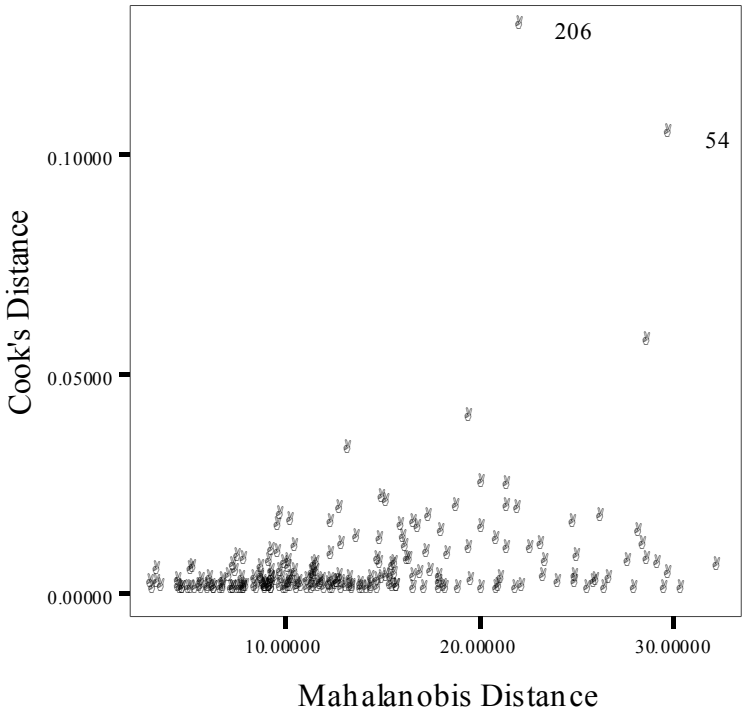

Figure 8. Scatter plot of Cook's to Mahalanobis distance for facets analysis.

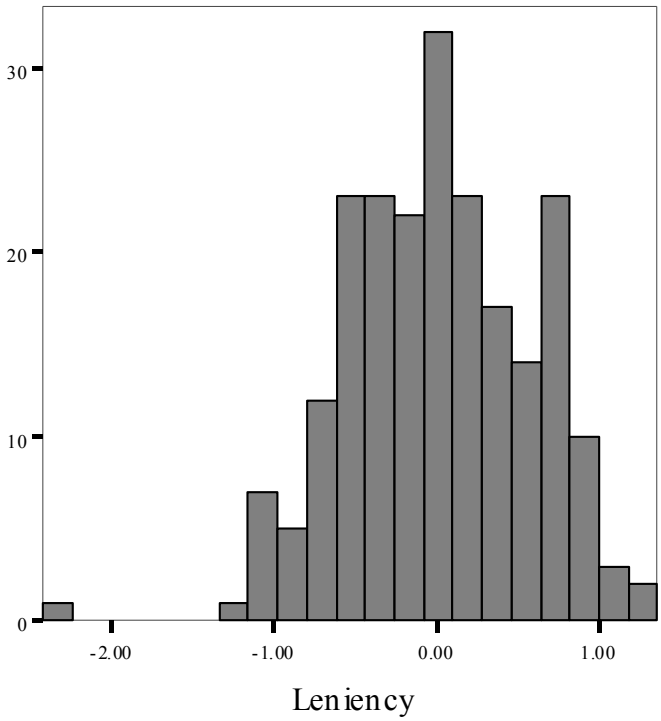

Figure 9. Histogram for rater leniency.

REFERENCES

Antonioni, D. & Park, H. (2001). The relationship between rater affect and three sources of 360 degree feedback ratings. *Journal of Management, 27,* 479-495.

Ashton, M. C. (1998). Personality and job performance: The importance of narrow traits. *Journal of Organizational Behavior, 19,* 289-303.

Ashton, M. C., Jackson, D. N., Paunonen, S. V., Helmes, E., & Rothstein, M. G. (1995). The criterion validity of broad factor scales versus specific facet scales. *Journal of Research in Personality, 29,* 432-442.

Athey, T. R., & McIntyre, R. M. (1987). Effect of rater training on rater accuracy: Levels-of-processing theory and social facilitation theory perspectives. *Journal of Applied Psychology, 72,* 567-572.

Bacon, D. R., & Novotny, J. (2002). Exploring achievement striving as a moderator of the grading leniency effect. *Journal of Marketing Education, 24,* 4-14.

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *Personality and Performance, 9,* 9-30.

Bartels, L. K., & Doverspike, D. (1997). Assessing the assessor: The relationship of assessor personality to leniency in assessment center ratings. *Journal of Social Behavior and Personality, 12,* 179-190.

Bass, B. M. (1956) Reducing leniency in merit ratings. *Personnel Psychology, 9,* 359-369.

Bernardin, H. J. (1980). The effect of reciprocal leniency on the relationship between consideration scores from the LBDQ and performance ratings. *Academy of Management Proceedings,* 131-135.

Bernardin, H. J. (1989). Increasing the accuracy of performance measurement: A proposed solution to erroneous attributions. *Human Resource Planning, 12,* 239-250.

Bernardin, Cooke & Villanova (2000). Conscientiousness and agreeableness as predictors of rating leniency. *Journal of Applied Psychology, 85,* 232-234.

Bernardin, H. J., Orban, J. A., & Carlyle, J. J. (1981). Performance rating as a function of trust in appraisal and rater individual differences. *Academy of Management Proceedings,* 311-315.

Bono, J. E., & Judge, T. A. (2004). Personality and transformational and transactional leadership. *Journal of Applied Psychology, 89,* 901-910.

Borman, W.C. (1977). Consistency of rating accuracy and rating errors in the judgment of human performance. *Organizational Behavior and Human Performance, 20,* 238-252.

Borman, W. C. (2004). Introduction to the special issue: Personality and the prediction of job performance: More than the big five. *Human Performance, 17,* 267-269.

Borman, W. C., & Hallam, G. L. (1991). Observation accuracy for assessors of work-sample performance: Consistency across task and individual-differences correlates. *Journal of Applied Psychology, 76,* 11-18.

Bretz, D. B., Milkovich, G. T., & Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions, and implications. *Journal of Management, 18,* 321-352.

Caldwell, D. F., & Burger, J. M. (1997). Personality and social influence strategies in the workplace. *Personality and Social Psychology Bulletin,* 1003-1012.

Cleveland, J. N., Murphy, K. R., & Williams, R. E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. *Journal of Applied Psychology, 74,* 130-135.

Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin, 90,* 218-244.

Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Costa, P. T., & McCrae, R. R. (1992). *NEO PI-R: Professional manual.* Lutz, FL: Psychological Assessment Resources, Inc.

Costa, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment, 64,* 21-50.

Drory & Ben-Porat (1980) Leadership style and leniency bias in evaluation of employee's performance. *Psychological Reports, 46,* 735-739.

Dudlely, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology, 91,* 40-57.

Eder, R. W., Keaveny, T. J., McGann, A. F., & Beatty, R. W. (1969). Evaluating faculty performance: An empirical investigation of factors affecting faculty ratings and student satisfaction using alternative rating formats. *Academy of Management Proceedings,* 23-27.

Farh, J. L., & Dobbins, G. H. (1989). Effects of self-esteem on leniency bias in self-reports of performance: A structural equation model analysis. *Personnel Psychology, 42,* 835-850.

Farh, J. L., Werbel, J. D. (1986). Effects of purpose of the appraisal and expectation of validity on self-appraisal leniency. *Journal of Applied Psychology, 71,* 527-529.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *Journal of Abnormal and Social Psychology, 44,* 329-344.

Fox, S., Ben-Nahum, Z., & Yinon, Y. (1989). Perceived similarity and accuracy of peer ratings. *Journal of Applied Psychology, 74,* 781-786.

Fox, S., Caspy, T. Reisler, A. (1994). Variables affecting leniency, halo and validity of self-appraisal. *Journal of Occupational and Organizational Psychology, 67,* 45-56.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology,* (Vol. 1, pp. 203-234). Hillsdale, CA: Sage.

Griffeth, R. W., & Bedeian, A. G. (1989). Employee performance evaluations: Effects of ratee age, rater age, and ratee gender. *Journal of Organizational Behavior, 10,* 83-90.

Guilford, J. P. (1954). *Psychometric methods.* New York, NY: McGraw-Hill.

Hammer, Kim, Baird, Bigoness (1973). Race and sex as determinants of ratings by potential employers in simulated work sampling task. *Journal of Applied Psychology, 59,* 705-711.

Harris, M. M., Smith, D. E., & Champagne, D. (1995) A field study of performance appraisal purpose: Research-versus administrative-based ratings. *Personnel Psychology, 48,* 151-160.

Hogan, J., & Roberts, B. W. (1996) Issues and non-issues in the fidelity-bandwidth trade-off. *Journal of Organizational Behavior, 17,* 627-637.

Holzback & Robert (1980) Psychometric properties of self-appraisals of job performance. *Personnel Psychology, 33,* 363-371.

Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods, 5,* 64-86.

Ilgen, D. R., & Favero, J. L. (1985). Limits to generalization from psychological research to performance appraisal processes. *Academy of Management Review, 10,* 311-321.

Jawahar, I. M. (2001). Attitudes, self-monitoring, and appraisal behaviors. *Journal of Applied Psychology, 86,* 875-883.

Jawahar & Stone (1997) Influence of raters' self-consciousness and appraisal purpose on leniency and accuracy of performance ratings. *Psychological Reports, 80,* 323-337.

Jawahar, I. M., & Williams, C. R. (1997). Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology, 50, 905-926.*

Jelley, R. B. (2005). Rater personality and performance-evaluation leniency. Unpublished doctoral dissertation, University of Western Ontario, Canada.

Jenkins, M., & Griffith, R. (2004). Using personality constructs to predict performance: Narrow or broad bandwidth. *Journal of Business and Psychology, 19,* 255-269.

John, O. P., & Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. In L.A. Pervin, & O.P. John (Eds.), *Handbook of personality: Theory and research.* (pp. 102-138) New York, NY: Guilford Press.

Johnson, T. P., & Fendrich, M. (2002). *A validation of the Crowne-Marlowe Social Desirability Scale.* Paper presented at the 57th Annual Meeting of the American Association for Public Opinion Research, St. Pete Beach, FL, May 2002.

Judge, T. A., & Ferris, G. R. (1993). Social context of performance evaluation decisions. *Academy of Management Journal, 36,* 80-105.

Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology, 87,* 530-541.

Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology, 87,* 797-807.

Kane, J. S. (1994). A model of volitional rating behavior. *Human Resource Management Review, 4,* 283-310.

Kane, J. S., Bernardin, H. J., Villanova, P., & Peyrefitte, J. (1995). Stability of rater leniency: Three studies. *Academy of Management Journal, 38,* 1036-1051.

Kingstrom, P. O., & Bass, A. R. (1981). A critical analysis of studies comparing behaviorally anchored rating scales (BARS) and other rating formats. *Personnel Psychology, 34,* 263-289.

Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin, 87,* 72-107.

Lee, Ashton & Shin (2005) Personality correlates of workplace anti-social behavior. *Applied Psychology: An International Review, 54,* 81-98.

LePine, J. A., Erez, A., & Johnson, D. E. (2002). The nature and dimensionality of organizational citizenship behavior: A critical review of meta-analysis. *Journal of Applied Psychology, 87,* 52-65.

Mandell (1956) Supervisory characteristics and ratings: A summary of recent research. *Personnel Psychology, 32,* 435-440.

McCoach, D. B., & Siegle, D. (2003). The School Attitude Assessment Survey-revised: A new instrument to identify academically able students who underachieve. *Educational and Psychological Measurement, 63,* 414-429.

McCrae, R. R., & Costa, P. T. (1999). A five-factor theory of personality. Theoretical perspectives. In L.A. Pervin, & O.P. John (Eds.), *Handbook of personality: Theory and research.* (pp. 139-153) New York, NY: Guilford Press.

Meyer, H. H. (1980). Self-appraisal of job performance. *Personnel Psychology, 33,* 291-295.

Miller, J. D., & Lynam, D. (2001). Structural models of personality and their relation to antisocial behavior: A meta-analytic review. *Criminology, 39,* 765-795.

Mount, M. K. (1984). Psychometric properties of subordinate ratings of managerial performance. *Personnel Psychology, 37,* 687-702.

Mount, M. K, & Barrick, M. R. (1998). Five reasons why the "Big Five" article has been frequently cited. *Personnel Psychology, 51,* 849-857.

Murphy, K. R., & Balzer, W. K. (1989). Rater errors and rating accuracy. *Journal of Applied Psychology, 74,* 619-624.

Murphy, K. R., & Dzieweczynski, J. L. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance, 18,* 343-357.

Norton, S. D., Gustafson, D. P., & Foster, C. E. (1977). Assessment for management potential: Scale design and development training effects and rater/ratee sex effects. *Academy of Management Journal, 20,* 117-131.

Nunnally, J. C. (1978). *Psychometric theory.* (2nd ed.) New York, NY: McGraw-Hill.

Paunonen, S. V. (1998). Hierarchical organization of personality and prediction of behavior. *Journal of Personality and Social Psychology, 74,* 538-556.

Paunonen, S. V., Rothstein, M. G., & Jackson, D. N. (1999). Narrow reasoning about the use of broad personality measures for personnel selection. *Journal of Organizational Behavior, 20,* 389-405.

Petty, R. E., Harkins, S. G., Williams, K. D., & Latane, B. (1977). The effects of group size on cognitive effort and evaluation. *Personality and Social Psychology Bulletin, 3,* 579-582.

Pulakus, E. D., Oppler, S. H., White, L. A., & Borman, W. C. (1989). Examination of race and sex effects on performance ratings. *Journal of Applied Psychology, 74,* 770-780.

Reardon & Waters (1979) Leniency and halo in student ratings of college instructors: A comparison of three rating procedures with implications for scale validity. *Educational and Psychological Measurement, 39,* 159-162.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology, 38,* 119-125.

Roach, D. W., & Gupta, N. (1992). A realistic simulation for assessing the relationships among components of rating accuracy. *Journal of Applied Psychology, 77,* 196-200.

Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin, 88,* 413-428.

Sauser, W. I., & Pond, S. B. (1981). Effects of rater training and participation on cognitive complexity: An exploration of Schneier's cognitive reinterpretation. *Personnel Psychology, 34,* 563-577.

Schneier (1977) Operational utility and psychometric characteristics of behavioral expectation scales: A cognitive reinterpretation. *Journal of Applied Psychology, 62,* 541-548.

Schoorman, F. D. (1988). Escalation bias in performance appraisals: An unintended consequence of supervisor participation in hiring decisions. *Journal of Applied Psychology, 73,* 58-62.

Sharon, A. T., & Bartlett, C. J. (1969). Effect of instructional conditions in producing leniency on two types of rating scales. *Personnel Psychology, 22,* 251-263.

Steele, R. P., & Ovalle, N. K. (1984). Self-appraisal based upon supervisory feedback. *Personnel Psychology, 37,* 667-685.

Steiger, J. H., & Fouladi, R. T. (1997). R2 Program (Version 1.1) [Computer software and manual]. Retrieved April 2, 2006, from http://www.statpower.net/page5.html

Sundvik, L., & Lindeman, M. (1998). Acquaintanceship and the discrepancy between supervisor and self-assessments. *Journal of Social Behavior and Personality, 13,* 117-126.

Tagger, S., & Brown, T. C. (2006). Interpersonal affect and peer rating bias in teams. *Small Group Research, 37,* 86-111.

Tett, R. P., Steele, J. R., & Beauregard, R. S. (2003). Broad and narrow measures on both sides of the personality-job performance relationship. *Journal of Organizational Behavior, 24,* 335-356.

Thornton, G. C. (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology, 33,* 263-270.

Tsui, A. S., & Barry, B. (1986). Interpersonal affect and rating errors. *Academy of Management Journal, 29,* 586-599.

Turban, D. B., & Jones, A. P. (1988). Supervisor-subordinate similarity: Types, effects, and mechanisms *Journal of Applied Psychology, 73,* 228-234.

Tziner, A. (1984). A fairer examination of rating scales when used for performance appraisal in real organizational setting. *Journal of Occupational Behavior, 5,* 103-112.

Wagner, S. H., & Goffin, R. D. (1997). Differences in accuracy of absolute and comparative performance appraisal methods. *Organizational Behavior and Human Decisions Processes, 70,* 95-103.

Waldman & Thornton (1988) A field study of rating conditions and leniency in performance appraisal. *Psychological Reports, 63,* 835-840.

Wallace, J. C., & Vodanovich, S. J. (2003). Workplace safety performance: Conscientiousness, cognitive failure and their interaction. *Journal of Occupational Health Psychology, 8,* 316-327.

Wexley, K. N., & Youtz, M. A. (1985) Rater beliefs about others: Their effects on rating errors and rater accuracy. *Journal of Occupational Psychology, 58,* 265-275.

Williams, K. J., DeNisi, A. S., Meglino, B. M., & Cafferty, T. P. (1986). Initial decisions and subsequent performance ratings. *Journal of Applied Psychology, 71,* 189-195.

Wright, R. E., & Palmer, J.C. (2006). A comparative analysis of different models explaining the relationship between instructor ratings and expected student grades. *Educational Research Quarterly, 30,* 3-18.

Yun, G. J., Donahue, L. M., Dudley, N. M., & McFarland, L. A. (2005). Rater personality, rating format, and social context: Implications for performance appraisal ratings. *International Journal of Selection and Assessment, 13,* 97-107.

Zhao, H., & Seibert, S. E. (2006). The big five personality dimensions and entrepreneurial status: A meta-analytical review. *Journal of Applied Psychology, 91, 259-271.*